

# **For Reference**

---

**NOT TO BE TAKEN FROM THIS ROOM**



Ex LIBRIS  
UNIVERSITATIS  
ALBERTAENSIS





Digitized by the Internet Archive  
in 2022 with funding from  
University of Alberta Library

<https://archive.org/details/Squire1978>





THE UNIVERSITY OF ALBERTA

PROBABILITY CONCEPTS IN GRADES ONE, TWO, AND THREE

by



BARRY FREDERICK SQUIRE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF ELEMENTARY EDUCATION

EDMONTON, ALBERTA

FALL, 1978





## ABSTRACT

Research has shown that it is feasible to teach probability in upper elementary grades and that younger children appear to have some grasp of several basic ideas about probability prior to formal instruction.

The purposes of the present study were: (1) to determine the status of six basic probability concepts in grade one, two, and three children, (2) to investigate the level of quantification of probability present in these subjects, (3) to investigate differences in response due to the embodiment of probability settings, and (4) to examine the effect upon response due to the factors sex, grade, and IQ.

The six concepts studied were: events in a sample space, the most favorable event, the most favorable sample space for a given event, sample space equally favorable to a given set of events, impossible event, and certain event.

Seventy-two grade one, two, and three students from a suburban school were tested individually in an interview situation. Subjects made choice responses on concept items and predictions on quantification items.

Results showed that four of the concepts were understood by at least 75% of subjects in each grade and scores on all concept items were significantly greater than is attributable to chance. The scores on the quantification items were lower in all grades with an overall average of 42% correct on these items. For grades one and two the numbers of correct responses





were not significantly better than chance. There were no significant differences in scores due to different probability settings but performance decreased as the number of trials in the sample space increased.

Each concept question was presented in three embodiments, spinner, block, and box. No significant effect was found due to embodiment and there were no interactions between embodiment and sex or IQ.

On the total test score sex, grade, and IQ were all found to have significant effects. Grade and IQ were significant factors in the concept scores but no significant effects were found in the quantification scores. No interactions were found between any of the factors on the criterion measures.

Subjects were asked to state reasons for their responses in the probability test. Many correct rationalizations were given for concept responses but very few in relation to quantification predictions.

This study found that there was a substantial increase in understanding of probability concepts in children as they pass through grade three. At the same time there was little understanding of the numerical relationships between probability settings and frequencies of the outcomes, except when the number of outcomes is small.

Several implications are drawn for teachers of lower grades and for curriculum writers. It is suggested that provision be made for informal consolidation of existing concepts and ideas for grades one and two and a wider range of activities and





experiences from grade three onwards in which students are led to further concepts and into quantification of probability.



## ACKNOWLEDGEMENTS

The writer is grateful for all the assistance and cooperation he received in conducting and reporting the study. In particular he wishes to express his sincere appreciation to the following:

Dr. W. G. Cathcart, his supervisor, for his help, encouragement, special interest, suggestions, and criticisms throughout the process of conducting the study and writing this report.

Dr. J. Kirkpatrick, a member of the thesis supervisory committee and earlier the writer's program advisor, for her continuing interest, and particularly her assistance in reading and commenting on the original draft of the report.

Dr. S. Hunka, an interdepartmental member of the thesis committee, for his interest in the study, for his advice in the analysis of the data, and for his consent to serve as a thesis committee member.

Dr. D. Sawada for his consent to serve as a thesis committee member and for his advice on various occasions throughout the study.

Gratitude is also expressed to the principal and teachers of the school in which the data were collected for their complete cooperation. Appreciation is also expressed to the Edmonton Public School Board for permission to conduct the study in the school.

A special word of thanks goes to the writer's wife, Jean, for her assistance, patience, and encouragement throughout the study and for typing the draft and final forms of the report.

B. F. S.





## TABLE OF CONTENTS

| CHAPTER                                                | PAGE |
|--------------------------------------------------------|------|
| I. INTRODUCTION AND STATEMENT OF THE PROBLEM . . . . . | 1    |
| Introduction . . . . .                                 | 1    |
| Statement of the Problem . . . . .                     | 6    |
| Major Questions and Hypotheses . . . . .               | 8    |
| Definition of Terms . . . . .                          | 10   |
| Delimitations and Limitations . . . . .                | 12   |
| Assumptions . . . . .                                  | 13   |
| Significance of the Study . . . . .                    | 14   |
| Organization of the Rest of the Report . . . . .       | 16   |
| 2. REVIEW OF RELATED LITERATURE . . . . .              | 17   |
| Piaget and Related Studies . . . . .                   | 18   |
| The Theory . . . . .                                   | 18   |
| The Experiments . . . . .                              | 19   |
| Criticisms and Related Studies . . . . .               | 26   |
| The Status of the Concepts and                         |      |
| Significant Factors . . . . .                          | 29   |
| Classroom Studies of Materials and                     |      |
| Instructional Methods . . . . .                        | 35   |
| Summary . . . . .                                      | 45   |
| 3. INSTRUMENTATION AND RESEARCH PROCEDURES . . . . .   | 48   |
| Instrumentation . . . . .                              | 48   |
| Probability Test . . . . .                             | 48   |
| Description of the Test Items . . . . .                | 53   |
| Pilot Study of the Probability Test . . . . .          | 60   |
| Canadian Cognitive Abilities Test . . . . .            | 61   |





| CHAPTER                                          | PAGE |
|--------------------------------------------------|------|
| Research Procedures . . . . .                    | 62   |
| Selection of Sample . . . . .                    | 62   |
| Data Collection . . . . .                        | 63   |
| Analysis of the Probability Instrument . . . . . | 65   |
| Reliability . . . . .                            | 70   |
| 4. RESULTS OF THE INVESTIGATION . . . . .        | 71   |
| Status of the Probability Concepts . . . . .     | 71   |
| Level of Quantification of Probability . . . . . | 74   |
| Effect of Embodiment . . . . .                   | 77   |
| Effects of Sex, Grade, and IQ . . . . .          | 80   |
| Rationalization Used by Subjects . . . . .       | 86   |
| Summary . . . . .                                | 91   |
| 5. SUMMARY, DISCUSSION, IMPLICATIONS,            |      |
| AND RECOMMENDATIONS . . . . .                    | 94   |
| Summary of the Investigation . . . . .           | 94   |
| Discussion of the Findings . . . . .             | 97   |
| Some Implications of the Findings . . . . .      | 104  |
| Recommendations for Further Research . . . . .   | 105  |
| * * *                                            |      |
| BIBLIOGRAPHY . . . . .                           | 108  |
| APPENDIX A GAME BOARD USED IN THE TEST . . . . . | 114  |
| APPENDIX B FREQUENCIES AND RELATIVE FREQUENCIES  |      |
| OF QUANTIFICATION RESPONSES . . . . .            | 116  |



# LIST OF TABLES

| Table | Description                                                                                                | Page |
|-------|------------------------------------------------------------------------------------------------------------|------|
| 1.    | Proportions of Blue Red, Yellow in the<br>Devices used in the Probability Test . . . . .                   | 49   |
| 2.    | Items and Embodiments by which Concepts<br>and Quantification were Tested . . . . .                        | 52   |
| 3.    | Mean and Standard Deviation on Probability<br>Test and Subtests using Total Sample and<br>Grades . . . . . | 66   |
| 4.    | Probabilities for Scheffe Multiple<br>Comparisons of Grade Means on Total<br>and Subtest Scores . . . . .  | 67   |
| 5.    | Skewness, Kurtosis, and Chi-Square for<br>Test and Subtests on Total Sample . . . . .                      | 67   |
| 6.    | Percentages Correct on Probability Test<br>Items for Total Sample . . . . .                                | 69   |
| 7.    | Test-Retest Reliability of Probability<br>Test . . . . .                                                   | 70   |
| 8.    | Mean Per Cent Correct Responses for Probability<br>Concepts in Each Grade and the Whole Sample . . . .     | 72   |
| 9.    | Mean Per Cent Correct Responses on Quantification<br>Items in Each Grade and the Whole Sample . . . . .    | 75   |
| 10.   | Comparison of Embodiment Means for Each Grade<br>and the Total Sample . . . . .                            | 78   |





| Table | Description                                                                                                    | Page |
|-------|----------------------------------------------------------------------------------------------------------------|------|
| 11.   | Summary of Three Analyses of Variance<br>Within Subjects Due to Embodiment and<br>Sex, Grade, and IQ . . . . . | 79   |
| 12.   | Summary of Analysis of Variance for<br>Responses to the Test and Subtests . . . . .                            | 81   |
| 13.   | Criterion Means by Sex, Grade, and IQ . . . . .                                                                | 82   |
| 14.   | Criterion Means by Sex and IQ Nested<br>Within Grades . . . . .                                                | 83   |
| 15.   | Frequencies of Rationalization Responses<br>for Concepts by Grades and for the<br>Total Sample . . . . .       | 88   |
| 16.   | Rank of Embodiment Responses Within Grades . . . . .                                                           | 102  |
| 17.   | Frequencies and Relative Frequencies of<br>Responses on the Six-Trial Quantification<br>Questions . . . . .    | 117  |
| 18.   | Frequencies and Relative Frequencies of Responses<br>on the Twelve-Trial Quantification<br>Questions . . . . . | 118  |



LIST OF FIGURES

| Figure | Description                                                                            | Page |
|--------|----------------------------------------------------------------------------------------|------|
| 1.     | Distribution of Scores on the Probability<br>Test by Grades and Total Sample . . . . . | 68   |
| 2.     | Total Scores by IQ-Sex Groups for<br>Each Grade . . . . .                              | 84   |
| 3.     | Concept Scores by IQ-Sex Groups<br>for Each Grade . . . . .                            | 85   |
| 4.     | Quantification Scores by IQ-Sex Groups<br>for Each Grade . . . . .                     | 85   |





## CHAPTER 1

### INTRODUCTION AND STATEMENT OF THE PROBLEM

#### I. INTRODUCTION

One of the main tasks of curriculum design is the selection of appropriate content material. In recent years the criterion for judging topics to be appropriate for school age children has tended to be how well the study of such topics prepares those children for meeting life's situations. Most parents and educators have regarded a thorough grounding in the basic skills of language, expression, and arithmetic as providing this preparation yet it has been difficult to find agreement as to precisely what basic skills are necessary to cope with life.

Within the field of elementary school mathematics, there are many skills to be mastered and ideas to be understood, yet one topic has consistently been omitted from prescribed curricula even though suggested for inclusion by many writers in the past twenty years. This topic is the study of probability. The need for all children to have some experience in ideas of probability was well expressed by Cohen (1957) in a perceptive comment about the nature of schooling. Cohen said that

Our system of education tends to give children the impression that every question has a single definite answer. This is unfortunate, because the problems they will encounter in later life will generally have an indefinite character. It seems important that during their years of schooling children should be trained to recognize degrees of uncertainty.  
(p. 137)



Restle (1961), Cohen (1964), and Estes (1976) reiterate how important a role subjective probability judgments play in our lives. We explain decisions made and conclusions reached on the basis of what likelihood we assign to events that are at best uncertain. Racha-Intra (1977) emphasized this function when he described the primary purpose of teaching probability as providing "a tool by which students comprehend the uncertainty model of the world" (p. 2). Yee (1966) called for the study of probability in elementary school on the grounds of the need to train children in decision-making skills.

The report of the Cambridge Conference on School Mathematics in 1963 recommended strongly that probability be a vital and appropriate part of the elementary school mathematics program. The National Advisory Committee on Mathematical Education (NACOME) (1975) similarly reported widespread support for the inclusion of probability as a necessary component of the elementary school curriculum. At the same time, the NACOME report cited a National Council of Teachers of Mathematics survey which found a deficiency in the training of teachers in probability and statistics resulting in a minimal treatment of probability topics by teachers in school. The instruction at school was found to be generally restricted to traditional graphing exercises and elementary descriptive statistics.

Although Fischbein (1976) observed "an increasing tendency to introduce the theory of probability in mathematical curricula" (p. 23), there has been a noticeable absence of this topic from



the official published curricula of most North American, British, and Australian elementary schools. At the same time there is no shortage of published programs and textual material on this topic designed for use in elementary schools. Examples are the School Mathematics Study Group series of four texts and commentaries, the Nuffield Mathematics Project Sequences on graphing and probability culminating in the "weaving guide" unit on probability and statistics, the Scottish Mathematics Project Series with relevant chapters for grades four through six, and sundry modern school texts which devote a chapter or two to activities involving chance and probability concepts.

The situation at present is that probability is included in some elementary school programs but not in others, the basis for inclusion often being the experience of the teacher. Three things need to happen before one can feel confident that all children will experience appropriate instruction in probabilistic ideas:

1. Starting points need to be determined by ascertaining what young children know about probability before receiving any formal instruction on the topic.

2. Curriculum materials then need to be prepared which begin at those starting points. Some of the already published programs and texts may be usable here.

3. Teachers and trainee teachers need inservice and pre-service courses to prepare them for teaching the material at appropriate levels.

It is the first of these tasks that this study was





designed to explore. The Cambridge Conference Report (1963) and Ausubel's (1968) theory of learning both had some influence on the design of the investigation.

The Cambridge Report called for as early a start as possible on the topic of probability; therefore grade one, two, and three children were chosen for this study. Most earlier studies involved pre-school and higher grade pupils with only one study investigating the first three grades.

It is assumed that meaningful learning is the desired product of classroom experiences. This, according to Ausubel (1968), occurs when two conditions are satisfied; the learner's anchoring ideas are ascertained, and the new material being presented is organized and modified accordingly. This agrees with Bruner's (1960) belief that subject matter can be modified and passed through stages of readiness just as children's thinking processes are. Ausubel goes further and says the subject matter must be modified and organized to match the child's readiness. The secret of teaching, he says, is to ascertain what the learner already knows and teach him accordingly. It seemed appropriate then to find the state of readiness of young children with respect to probability concepts prior to organizing learning experiences for them.

Reasons for the present study also came from the research literature. Many studies have been concerned with the level of understanding of chance at various ages and what conditions or environmental factors appeared to influence the judgments



made by children at these ages. The first conclusion to note is that "the global intuitions of relative frequency and probability are present even in pre-operational children" (Fischbein, 1976, p. 29). At this level, and for concrete operational children, many researchers have found that children have the ability to make only comparisons that are reducible to a binary operation. Not until the formal operational stage, according to Fischbein's interpretation of Piaget, can children make the synthesis between the possible and the deductive and understand quantification of probability as the relationship between the number of favorable and possible outcomes. This conclusion may have in fact contributed largely to the lack of research on probability at the elementary school level.

There have been those who claimed that preoperational children's comparison judgments have to be interpreted only as perceptual comparisons and not as probability judgments (Hoemann & Ross, 1971). Fischbein (1976) makes the point that probabilistic judgment is not completely reducible to quantitative estimation. There is a large element of understanding the meaning of the situation which implies an "intuition of chance" (p. 36). The existence of such an intuition appears to be indicated by Strohner and Nelson's (1974) finding that preoperational children are intuitively able to distinguish between probable and improbable events. In situations where there is a conflict between verbal meaning and factual probability the latter is stronger. For example, a 3-year-old child produces the situation "the girl feeds the baby" when asked to show "the baby





feeds the girl", as the former is seen to be more probable than what is requested.

Fischbein (1976) also holds the view that the intuitive background of probabilistic thinking desired in children needs to be built. If the spontaneous background of probabilistic thinking found in children is a mixture of correct and incorrect intuitions, as he suggests, then further research needs to be done to investigate the mixture at all ages up to adolescence so as to determine the foundation upon which curriculum building can occur.

Other research has found that the responses of young children tend to be egocentric, predictions are affected by an alternation tendency, reward is a significant factor at all ages, and mode of representation affects response. A number of investigations also found that pre-school children can "learn about probability" if the conditions tend to be largely nonverbal and reinforcing.

The present study sought to add to the current knowledge about early probabilistic thinking for the reasons outlined above. The study sought to embody several suggestions made by previous researchers as well as utilize features of past designs.

## II. STATEMENT OF THE PROBLEM

The purposes of the present study were: (1) to determine the status of six basic probability concepts in grade one, two, and three children, (2) to investigate the level of quantification of probability achieved by these subjects, (3) to investigate



differences in response due to the embodiment of the probability setting, and (4) to examine the effect on performance on probability tasks due to the factors sex, grade, and IQ.

The six probability concepts in question were:

1. events in a sample space,
2. the most favorable event in a sample space,
3. the most favorable sample space for a given event,
4. sample space equally favorable to a given set of events,
5. impossible event, and
6. certain event.

These are defined in a later section of this chapter.

With respect to the second purpose there is some dispute among researchers as to when children begin to attend to the quantitative aspects of a probability setting. Fischbein, Pampu, and Manzat (1970) and Chapman (1975) found young children able to correctly compare ratios of the form  $a/b:c/b$  where the comparison could be reduced to a binary operation,  $a:c$  in this case. Fischbein et al. also found that 9-year-olds quite readily made correct probability judgments on the basis of relative frequencies of preferred events. On the other hand Piaget (1975) maintained that proportional facility in probability situations only comes with the formal operations stage of development. It was hoped that this study would add to the evidence on young children's ability with quantification of probability.

The third purpose, the effect of embodiment on pupil response, was investigated partly to check Jones' (1975) findings and



partly to provide further guidance for the design of activities and other curriculum material.

Wilkinson and Nelson (1966) found that responses were affected by the familiarity of the situation to the subject. The devices and situations used in the present study were designed with this finding in mind. Every attempt was made to make all embodiments a new experience for the subjects.

The consideration of individual differences is vital for effective teaching. Sex, grade, and IQ are three easily discernible differences among pupils, thus the effect of these factors on response to probability questions was examined as the fourth purpose of this study. No attempt was made to test for the effect of the many other social and personal characteristics such as socio-economic status, cognitive style, conservation ability, vocabulary, and listening ability. Some of these may have contributed to error in the criterion measure and would thus impose a limitation on the study. Many of them would be worth examining in a further investigation.

### III. MAJOR QUESTIONS AND HYPOTHESES

The four purposes stated above gave rise to the following questions and null hypotheses.

#### Purpose One

1. What proportion of subjects in each grade and in the total sample indicate an understanding of the six concepts investigated? These relative frequencies are taken to be





measures of the status of the concepts.

2. Null Hypothesis: The proportions derived in answering question one are not significantly different from chance proportions for each concept.

#### Purpose Two

3. What proportion of subjects in each grade and in the total sample indicate an understanding of the quantification items presented?

4. Null Hypothesis: The proportions derived in question three do not differ significantly from chance proportions.

#### Purpose Three

##### Null Hypotheses.

5. There are no significant main effects due to embodiment on performance on the probability test.

6. There are no significant main effects due to embodiment on performance on the probability test when sex, grade, and IQ are used as blocking variables in pairs.

7. There are no significant interactions between the embodiment and the factors sex, grade, and IQ.

#### Purpose Four

##### Null Hypotheses.

8. There is no significant main effect on probability test performance due to (a) sex, (b) grade, and (c) IQ.

9. There are no significant interactions between the independent variables sex, grade, and IQ on the criterion measures.



#### IV. DEFINITION OF TERMS

For the purposes of this study and report, the following terms are used as defined.

##### Probability Concepts

Probability concepts are taken to be the notions or ideas about random phenomena such as the outcome of rolling a die or whirling a spinner. The concern is more with "what will happen?" rather than "how often?".

##### Quantification of Probability

Quantification of the probability of an event is the assigning of a relative frequency to that event. This is a predicted measure of "how often" that event can be expected to occur. Due to the age of the subjects in the study proportions and fractions were avoided and expressions such as "four out of six" were used.

##### Embodiment of Probability

Situations were presented in three modes or embodiments: a spinner, a block, and a box containing counters of various colors. All had six "outcomes" to enable the construction of equivalent probability settings.

##### Random Devices

The devices referred to occasionally are the spinners, blocks, and boxes arranged as outlined in chapter 3 under Materials.



### Probability Setting: a-b-c

Throughout the investigation, five different trinomial proportions were represented by the devices. These are referred to as settings. For convenience of labelling, the proportion  $a:b:c$  is rewritten a-b-c and used to identify which device is being used. For example, the 2-3-1 spinner means the spinner with two BLUE, three RED, and one YELLOW segment. This color order was invariant throughout the experiment.

### Events in a Sample Space

A listing of all possible outcomes in a given situation constitutes the events in a sample space.

### The Most Favorable Event in a Sample Space

This is the event with the greatest expected chance or relative frequency of occurring. For example, in using the 2-3-1 spinner RED would be the most favorable event.

### The Most Favorable Sample Space for a Given Event

This is the sample space in which the given event has the greatest chance of occurring. For example, given the 2-2-2, 3-2-1, and 4-1-1 blocks the most favorable sample space for BLUE to occur is the 4-1-1 setting.

### Equally Favorable Sample Space

This is the setting in which all events have the same chance of happening. In the previous example the 2-2-2 setting gives all events (colors) an equal chance of occurring on any one roll of the block.





### Impossible Event

This is a phenomenon which is not in the sample space and hence cannot occur. For example, with the 3-3-0 spinner YELLOW would be an impossible event.

### Certain Event

This is a phenomenon which is the only element in the sample space being considered and hence must occur every time. For example, with the 0-0-6 box YELLOW would be a certain event on each draw as the box contains only yellow counters.

### Understanding

Understanding of a concept or a principle is indicated by the average per cent of correct responses on items relating to that concept or principle.

## V. DELIMITATIONS AND LIMITATIONS

The sample for the study was selected from grade one, two, and three children at the one school in Edmonton made available by the Edmonton Public School Board. The subjects were selected on the basis of teacher perception of general reasoning ability within grades and sexes. The opportunity did not exist for a more random sample from a larger population.

The delimitations stated above impose some limitations on the generalizability of the results. The grade three classes tended to have very few low-IQ girls from which to select, and this would not have been representative of the wider population. Teacher perception is a subjective judgment, but in this case



a correlation of 0.778 with IQ scores was found across the sample when IQ scores were later available for grade one and two subjects.

No attempt has been made to control for socio-economic status, although an average rating was calculated for the overall sample. Rural students may also have different academic and social experiences from urban students and may respond differently to the test items and indicate a different level of understanding of probability concepts. It will be left to further research to determine if the differences mentioned above do exist.

## VI. ASSUMPTIONS

One major assumption of the present study was that subjects' choice responses given in the test provided a true picture of the level of understanding of the related probability concepts. This was an assumption about both the validity of the instrument and about the child's understanding of the questions asked. A panel of experts judged the instrument as being a valid test of the basic concepts in question. The interview protocol was designed to ensure that subjects understood the questions asked.

A related assumption that is unavoidable in all forms of verbal testing is that the concepts and explanations expressed were the concepts and reasons believed. The researcher was successful in winning the confidence of the children so that most of them appeared to be relaxed enough to respond freely and openly. There is always that uncertainty, however, that accompanies one's judgment in such situations.

It was assumed that no subjects had received any formal



instruction in probability concepts. The school program contained no such activities up to the grade four level. It could not be assumed, on the other hand, that pupils had had no experience with notions of chance, as most acknowledged that they had played games of chance involving dice and/or spinners. This would be difficult to control for.

It was also assumed that no learning took place from one item on the test to the next. Students were not told whether they were correct or not, but were rather given a non-committal "O.K.", "Uh-uh", or a repeat of their answer with an inflexion. In addition, possible teaching effect was controlled for by randomizing the order of presentation of the items as described in detail in chapter 3. If learning did take place on an individual level it would then be uniformly distributed among the items. The interviews took only about twenty minutes each.

Finally, it was assumed that the IQ ratings were reliable measures of comparative cognitive abilities within the sample. The test writers advise that the ratings only be used on a comparative basis and within three months of the test being administered. Both these conditions were satisfied.

## VII. SIGNIFICANCE OF THE STUDY

One of the points made in the introduction to this report was that educators' long-standing recommendations for early study of probability at school have not been implemented. Very few curricula include the study of probability as part of a core of study for all elementary school students.





Following the Cambridge Report's recommendation that the study of probability begin as early as possible in the elementary school, several research studies have been concerned with the design and teaching of instructional units in probability. Wilkinson and Nelson (1966) found that grade six children responded well to a specially designed unit but most subjects had strong preconceived intuitions about situations familiar to them from earlier experience. Many intuitions that were incorrect proved difficult to alter. Ojemann, Maxey, and Snider (1965, 1966) found that preliminary instruction of grade three children has a positive effect on probabilistic behavior. Fischbein, Pampu, and Manzat (1970) demonstrated that even 10-year-old children can rapidly understand the concepts of arrangements and permutations and solve problems involving simple combinatorial calculations. Shepler (1969) found it was possible to teach introductory probability and statistics to sixth-grade students to a high degree of mastery but he cautions against the use of activities and materials that call for too subtle an interpretation.

All the researchers indicate a belief that probability should be taught in elementary schools and call for further investigation into what should be taught, when experiences in probability should begin, in what sequence, and with what rigor. As indicated in the introduction, this study is concerned with finding anchorage points for the development of sequences of probability activities for young children. If meaningful learning experiences can then be organized into the curricula of the lower elementary grades, we may be able to prevent many



incorrect intuitions of chance developing and lay a more solid foundation for later studies of probability and statistics.

#### VIII. ORGANIZATION OF THE REST OF THE REPORT

Chapter 2 contains a review of the literature relevant to the development of notions of chance in children. Chapter 3 presents a description of the probability test and its design, the IQ test, research procedures, and analysis of the probability instrument. Chapter 4 presents the results of the data analysis and chapter 5 gives a summary and discussion of the main findings.



## CHAPTER 2

### REVIEW OF RELATED LITERATURE

Much of the related research has been concerned with the feasibility of teaching concepts of probability in the upper elementary grades or later. This may well have been influenced by Piaget's statement that the quantification of probability demands the onset of formal operational thought. Whatever the reason, we know very little about the actual readiness of young children for this topic.

The research that has been done can be divided into three main groups:

1. Piaget and Inhelder's study of the child's understanding of chance (1975), and studies directly related to their resulting theory of development of the idea of chance,
2. studies concerned with the status of probability concepts and significant factors affecting their development, and
3. studies concerned primarily with the preparation and trial of instructional units in probability and their implementation with, and effects upon, elementary school children.

There is unavoidable overlap between these categories as some studies which began with Piaget's theory also examined factors affecting the development of chance concepts. Likewise, others could have been reported in either group 2 or group 3. As near as is practical, the studies are reported in chronological order within each section.





## I. PIAGET AND RELATED STUDIES

### The Theory

Piaget's (1960) genetic epistemological theory of development is well known among researchers for the way it presents intellectual development as the organization of intellectual operations into structured systems, progressively producing cognition of greater genetic maturity. Logical and arithmetical operations are seen as internalized actions organized into systems which are basically characterized by rigorous composition and reversibility. Such reversible composition makes possible deduction and lawful predictions. However, chance transformations and chance events are not rigorously reversible and cannot be rigorously composed in deductive systems.

Piaget (1975) applied his general theory to the development of the concept of chance in children and proposed the following:

1. The preoperational child is influenced more by contiguity in space and time than by causality, and he is unable to distinguish possibility from necessity. Fischbein (1976) elaborates on Piaget's meaning: "Being unable to understand necessity, he is unable to understand-by-opposition the unnecessary, the fortuitous" (p. 27). In the words of Flavell (1963), "for the preoperational child, nothing is deductively certain and nothing is genuinely fortuitous . . . ; his thought is forever at midstation between these poles" (p. 342).

2. During the concrete operational stage (approximately age 7 through age 10) the child begins to separate the necessary



from that which is simply possible; he becomes able to distinguish chance events from the strictly deductive ones. This is not enough to produce probabilistic thinking in its full sense because the child only understands the possibility of operational systems related to chance events as he discovers them in an incomplete and empirical fashion.

3. Not until the formal operational stage (about age 12) can the probability concept begin to be completely mastered. Ability with combinatoric operations and proportions at this stage allows for the synthesis between the possible and the deductive which is the source of probabilistic thinking.

### The Experiments

Except for a brief article in 1950, Piaget's only work on chance concepts in children is the book he and Inhelder wrote in 1951 which was not translated into English until 1975. The conclusions given above arose from a dozen or so experiments reported in the book. There were three main experiments in Part I and three more of relevance in Part II which are summarized below.

### Part I. Chance in Physical Reality

#### 1. Random mixture and irreversibility.

To test for notions of random mixture and irreversibility a rectangular tray was set up as a see-saw in which 2 groups of balls, 8 red and 8 white, separated by a divider, were arranged along its width. At each see-saw movement, the balls would



roll to the opposite end then return to the original end when the box is tipped back. The balls invariably collided with each other and returned into a different arrangement. Prior to each tipping, the subjects were asked to predict the arrangement of the balls when they return to the starting end. Questions were used such as: Will the red ones stay on one side and white ones on the other? Or will they mix up? In what proportions?

After several successive predictions and tippings, subjects were asked to predict the result of a large number of moves of the tray. Of particular interest was any prediction of a progressively random mixture, or of a final reordering to their original sides, the two extremes which Piaget encountered.

## 2. Centred and uniform distributions.

(a) To test for notions about centred distributions, five funnelled boxes were shown to the subjects one at a time. The first four were funnelled in the centre of the top, the fifth at the right side of the top. Boxes #1 and #5 had 2 bins at the bottom, #2 had 3 bins, #3 had 4 bins, and #4 had 18 bins below a network of nails as in a regular quincunx.

As each box was used, several marbles were dropped in the funnel one at a time with the subject predicting then explaining the outcome. Finally, about sixty were let go with the subject first giving a prediction then an explanation.

(b) For uniform distributions, small glass beads which did not roll too easily were dropped from a kind of trellis





sieve onto a sheet of grid paper and the subjects were asked about the chance of uniform distribution as a function of the greater or smaller number of beads dropped. (p. 27)

3. Constant relationship in conflict with fortuitous uniform distribution.

To find out how the mind of the child succeeds in dissociating what is due to chance from what is due to non-chance, a more "flamboyant" (Flavell, 1963) experiment was used. The apparatus was essentially a roulette wheel with an iron bar as a pointer. There were 16 equal sectors of 8 colors, opposite sectors being of like color, and the wheel spun "honestly" until a set of matchboxes was placed on its colors. The matchboxes looked identical but all contained wax in which were embedded pieces of metal of various kinds to make four sets of four boxes of different mass. Two boxes in one set contained magnets so that the spinner could be made to stop on a chosen color.

The initial inquiry was about the honest wheel, that is, about where the pointer would be likely to stop on a given spin, about the distribution of stops over a large number of spins, and so on. Once the child had noted the fortuitous distribution of stopping places, the matchboxes were introduced to give an unexpected constant element designed to throw off the predictions of the child. This assumed he saw the dispersion of the stopping points as uncertain. His reactions to and explanations for this new situation were recorded.



### Results of Part I Experiments

The analysis of each experiment contains extensive examples from the subjects' protocols followed by the experimenters' explanations and hypotheses. These examples of children's reasoning make this book one of the easier and more interesting of Piaget's to read. In each of the three studies, the preoperational children made the more interesting responses. They tended to impute a hidden lawfulness to the randomization process in all three experiments (e.g., "blue this time as it was red last time"), and held quasi-magical views of causality on occasion (e.g., "it would go to green if you concentrated hard enough"). Flavell (1963) made a delightful footnote to gambling-oriented readers on this point referring to preoperational tendencies! (p. 344)

Responses of older subjects indicated a gradual ontogenetic development of notions of irreversibility with respect to the order of the marbles in the see-saw tray. Likewise, the ability to predict distributional form was slow in developing. This is in keeping with Piaget's view that a grasp of "the law of large numbers" depends on an understanding of proportions, something not acquired in force until the formal operational period.

With regard to the wheel experiment, younger subjects inferred more predictability than was justified and were not surprised by the results of the dishonest spins. They considered them

not beyond the pale of the hodge-podge of quasi-magical causal relations already thought to be at work in the genuinely random turns. (Flavell 1963, p. 344)



Older children, on the other hand, sensed a trick and went on to discover the cause.

## Part II. Random Drawings: The Experiments and Their Results

### 1. Chance and "miracle".

The first experiment in this section is similar in one aspect to the third experiment in part I in that it allowed the experimenter to "cheat chance", to produce "the miracle" at will. Two sets of about 15 counters were used; each one in the first set bearing a circle on one side and a cross on the other, each in the second having crosses on both sides. Showing the subject the first set, the experimenter asked for a prediction for a single throw, then for the distribution of crosses and circles if all were thrown at once. Then the experimenter surreptitiously substituted the second false set, threw them all at once, and gauged the child's reactions.

The younger children again merely registered mild surprise whereas older ones suspected a trick and quickly turned a few counters over to confirm their suspicions. The startling thing is that even after the trick was revealed, the younger subjects were inclined to think that the same result could readily be reproduced using the true counters.

This experiment was repeated in a slightly different form using two sacks of marbles. The first had red and blue marbles in it, the second had only blue marbles. Again the trick was employed, reactions analysed, and the trick explained.

At the end of each test, without the subject knowing



whether the mixed or homogeneous set was being used, the experimenter threw one by one the trick counters, or drew out the marbles one by one from the trick sack (blue). The aim was to catch the moment the subject recognizes for sure that it is an experiment with the homogeneous elements, and to explore his reasoning. According to Piaget, "this last experiment often gives the surest index of the judgment of probabilities of which the child is capable" (1975, p. 97).

In both experiments the reasons of the younger subjects appeared to be egocentric and phenomenological and Piaget concluded that these children understood nothing about the notion of random mixture.

## 2. Random drawing of pairs.

The first of the experiments having a bearing on quantification involved four unequal sets of different colored counters which were mixed thoroughly in a bag (e.g., 15 yellow, 10 red, 7 green, and 3 blue). Identical sets of each color were left on the table as a memory aid. The child made successive drawings of pairs of the counters from the bag and was asked to predict the most probable pair before each drawing. No replacements were made and each pair drawn was placed on the table in front of the subject as he drew them. It was possible for him to know what remained in the bag, but no explanations of this were given.

The reported results again matched Piaget's earlier conclusions. Younger children predicted according to a variety of bases, among them color preference and the imputed lawfulness





of "taking turns". Concrete-operational children tended to base their forecasts on relative frequencies but to forget that each drawing changed the frequencies, and so failed to keep their estimates up to date as they continued to draw each pair. The oldest children tended to keep a running tally of the changing distribution and quantified the probability as a function of the counters left in the sack.

### 3. Quantification of probabilities.

In the final study summarized here, counters were used, some with a cross on the back, the others with no cross. The experimenter made up two collections of counters and showed them to the subject. All collections had a small number of counters, for example, 2 with crosses, 2 without, and 1 with and 2 without. When the subject had noted the makeup of each set, they were separately mixed up and placed blank face up on the table. The task was to judge whether there was a greater chance of drawing a "cross counter" from one set than from the other. A number of such problems were posed, some very simple (e.g., comparing a 2-2 set with a 0-4 set) and others more difficult (e.g., a 1-2 set and a 2-3 set).

The same three developmental stages appeared in the responses, according to Piaget. In the four to seven year group, there was the absence of any comparisons based on the proportions in play. Occasionally there was an intuitive comparison when striking disproportions were perceived. For those in the second stage, seven to eleven years, there was a beginning of quantification but with a consistent error: the prediction was solely on the



absolute number of counters with crosses in each set rather than in terms of the ratio of these to total counters in each set. The subjects compared favorable with unfavorable cases but did not construct a relationship between the favorable and the possible. This relationship appeared to be established in all of the eleven-years-and-over subjects: they tended to solve each problem by a calculation with fractions.

Other experiments were reported by Piaget and Inhelder where they investigated even further the quantification of probabilities and combinatoric operations, but the six outlined above are sufficient background for this review which is concerned more with Piaget's stage I and II subjects.

#### Criticisms and Related Studies

Very few details were given by Piaget about the size and nature of his samples, and this is one of the grounds on which he is criticized. It was noted by Flavell (1962) that Piaget's probability experiments also called for a high level of verbalization of the child's understanding of various technical aspects of mathematical probability. All we are told about the samples is that the age range was from 3 to 12 years and samples of 7 to 14 subjects were used. No attempt was reported to control for effects of age, sex, or other relevant variables, and no provision was made for analysing the results statistically.

The main area of disagreement with Piaget has been with his conclusion that children up to the age of about 7 show no evidence of being able to think probabilistically. Contrary



evidence began to come from the work of Messick and Solley (1957), Stevenson and Zigler (1958), Stevenson and Weir (1959), and Siegel and Andrews (1962) who examined reinforcement conditions and found young children tended to adopt maximizing strategies in probabilistic situations.

Yost, Siegel and Andrews (1962) listed several shortcomings of Piaget's technique, designed a model to overcome these, and compared it with a Piaget-type procedure. Their desire to minimize the verbal content of the procedure led them to a decision-making technique involving a choice between two transparent boxes of counters instead of a prediction with one box and a duplicate set of tokens displayed in a fixed, initial manner. They also sought to control for color preference and provided reinforcement other than knowledge of the outcome. The scores under the two conditions were different enough for Yost et al. (1962) to conclude that

Probability judgments made by 4- and 5-year-old children have been observed in two situations. In the situation in which controls are introduced, amount of reinforcement is increased, and an opportunity for non-verbal decision-making is presented, children tend to make correct responses significantly more frequently (p. 780).

Goldberg (1966) replicated this experiment, with some modifications, endorsed the above conclusions, and emphasized the importance of task conditions in observing the level of performance of four- and five-year-olds on probabilistic judgment tasks.

The Yost and Goldberg studies involved only 19 and 32 subjects respectively and both researchers called for further





investigation. Davies (1965) designed and administered non-verbal and verbal tests on probability concepts to 112 subjects, 8 male and 8 female at each of ages three through nine years. Her results supported Piaget's interpretation of the acquisition of this concept as a developmental phenomenon but there appeared to be evidence that preoperational children frequently behave according to event probabilities even before they can adequately verbalize the probability concept by which they are responding. No significant sex difference was found in the development of this concept and Davies concluded with a suggestion that age of appearance of non-verbal and verbal demonstrations of probability concepts be related in future studies to other variables such as MA or IQ, socio-economic status and educational level of the home, and patterns of child-rearing.

Offenbach (1964, 1965) also found evidence for probabilistic thinking in kindergarten children and this contention seems to be refuted only by Hoemann and Ross (1971). They argued that preschool children's probability judgments are really only perceptual comparisons of two arrays of objects and not judgments about likely or unlikely outcomes at all.

Fischbein (1976) disagreed with Hoemann and Ross and contended that understanding the meaning of the situation in which they placed their subjects implies necessarily the intuition of chance and that this intuition and the intuitive estimation of probabilities appear early in childhood. This would seem to be the state of this argument at the present time,



but many more aspects are considered in the next two sections of this review.

## II. THE STATUS OF THE CONCEPTS AND SIGNIFICANT FACTORS

Some research has studied the type and level of chance ideas which are acquired naturally, that is, without formal instruction. This is the first concern of this section of the review.

Doherty (1966) examined the status of four concepts of probability in a sample of 54 grade four, five, and six children. She concluded that the subjects had acquired naturally, from everyday experiences, considerable familiarity and ability with the four concepts: (1) the idea of a sample space, (2) the probability of a simple event in a sample space, (3) the probability of the union of non-overlapping events in the sample space, and (4) the idea of the difference between mutually independent events and mutually exclusive events. No significant difference was found in level of difficulty between concepts nor were sex or chronological age significant factors for this sample. Significant differences were found between ability levels, mental age levels, and mathematical and average achievement levels.

Leffin (1969) surveyed 528 children randomly selected from the grade four through seven population of the Wausau Public School System, Wisconsin. The sample was categorized on the basis of sex and three IQ ranges. Three concepts were examined: (1) points in a finite space, (2) probability of a simple event in



a finite sample space, and (3) the quantification of probability. (The labelling of this last item as a concept is questionable.) Three tests were administered to groups of subjects, one test for each concept. The overall mean performances were significantly different among IQ groups, sex, and grades. The most significant outcome reported was that the children demonstrated that they had acquired considerable knowledge about the three concepts in question, and that such knowledge must have developed as a result of their background, experience, and intuition. Both Doherty and Leffin called for an inclusion of probability topics in the elementary curriculum for intermediate grades, and for the development of methods and materials appropriate for the task.

Jones (1975) appears to be the only researcher who looks specifically at grades one, two, and three with respect to performance on concepts of probability. The five concepts were: (1) outcomes of a sample space, (2) most favorable event in a sample space, (3) most favorable sample space for a given event (same number of outcomes in each space), (4) sample space equally favorable to a given set of events, and (5) the most favorable sample space for a given event (same number of favorable outcomes in each space). A sample of 162 subjects was chosen on the basis of the grades and three IQ levels. Three embodiments were used: spinners with equal sectors comprised the first embodiment (unit), spinners with unequal sectors comprised the second (gross), and containers of discrete objects comprised the third (set). Five tests were administered



individually by the investigator in an interview situation. Corresponding test items for the three embodiments were isomorphic in a probability sense. Also half the items were contiguous and half the items had non-contiguous outcomes. Items from both sets were matched in a probability sense.

It was found that grade two and three children performed significantly better than grade one children on concepts one, four, and five, but there were no significant differences between grades two and three on any of the tests. Several differences were reported regarding the embodiment and contiguity factors. This led the investigator to conclude that some topics in probability should be introduced into the primary curriculum but the concepts which are included should be presented in as many different settings as possible.

The second concern of this section of the review is with studies which have primarily focused on factors which affect children's response in probabilistic situations.

Offenbach (1964) systematically studied the effect of reward and punishment on the learning behavior of 30 kindergarten and 30 grade four children. Using marbles and ten-cent toys as the prizes for correct guessing of the next card in a specially made deck, he found that the reward-punishment groups of subjects chose the more frequent event more often than the control group. The level of reward-punishment did not appear significant, in agreement with the findings of Brackhill, Kappy, and Starr (1962). The absence of probability matching behavior in the control group was at variance with





results reported by Siegel and Andrews (1962). Offenbach suggested the inconsistencies could be attributable to methodological differences, simultaneous presentation of stimulus in pairs by Siegel versus successive presentation by Offenbach. In a subsequent study Offenbach (1965) found little effect due to the method of stimulus presentation except that the simultaneous procedure made it easier for both age groups (K and 4) to respond on the basis of previous outcomes. In both studies Offenbach found a tendency for the fourth graders to try to find rules governing the occurrence of the events while the kindergarten children responded to the immediate situation in isolation. The older children appeared to be more aware of a possible sequential nature of the task. These and other intratask behaviors reveal age differences consistent with Piaget's stages of logical thinking but take issue with Piaget's belief that probabilistic thinking doesn't occur before age seven.

Mullenex (1968) group-tested a class in each of third, fourth, fifth, and sixth grades to determine the level of understanding of four probability concepts and to study the relationship of this understanding to the variables of sex, age, general ability, and basic skill in school subjects. None of the variables appeared to be relevant predictors of the criterion measure, but judgment was reserved in regard to sex, basic reading skill, arithmetic skill, and problem solving skill as measured by the Iowa Tests of Basic Skills. The investigator recommended individual testing and instruction



to more precisely determine the relationship between these variables.

Carlson (1969) used several Piagetian-type tests of varying difficulties to examine 160 eight- to eleven-year-old children for development of probabilistic thinking. Apart from supporting Piaget's conception of the development of probability concepts in children at this level, Carlson noted that age and socio-economic status were the most significant factors, level of general intelligence was less important, and no differences in performance were attributable to the subject's sex.

Fischbein, Pampu and Manzat (1970) found that age and instructional conditions were highly significant factors in children's responses in a ratio-comparison experiment related to concepts of chance. Sixty children from each of grades K, three, and six\* were seen individually and presented with 18 problems. Twenty subjects at each grade level were assigned to one of these instructional conditions each of which began with a guessing game concerned with black and white marbles. The interesting conclusion drawn by the investigators is that a little amount of instruction enabled the nine-year-old subjects to correctly estimate chance by comparing ratios whereas prior to the brief instruction the nine-year-olds' spontaneous responses differed little from those of the five-year-olds. This finding led the investigators to question Paiget's hypothesis about the proportionality concept only

---

\* American schools equivalent would be grades 1, 4, and 7.



coming at the formal operational stage and to argue that probability topics should be started in primary school.

Hoemann and Ross (1971) conducted four experiments on probability judgment tasks with children ranging in age from preschool to early adolescence. As mentioned in part I, their point of view was that the preoperational child made supposed probability choices only on the basis of magnitude discrimination with no probability inference being involved. They acknowledged that there was a contrary point of view but argued that their results supported Piaget and Inhelder's account of the development of probability concepts in children.

One particular result from Hoemann and Ross's study of relevance to the present study was that

preschool children of CA  $4\frac{1}{2}$  performed at the 75% level on the two-array task that allowed a direct comparison, while this level was not reached until CA 6 in a single-array task. (p. 235)

The single array task called for a prediction which Piaget had suggested was more difficult than a comparison choice, although his experiments required mostly prediction responses. Yost, Siegel, and Andrews (1962) and Goldberg (1966) had utilized choice between arrays instead of prediction with one array in their decision-making models and found improved response as a result.

Hoemann and Ross found it unnecessary to employ the elaborate controls for color preference and odds displays that Yost et al. and Goldberg had used. Their use of black and white is an indication of this but they also state the fact





explicitly. While acknowledging that preschool children do indeed sometimes prefer color to quantity, they affirm that no probability inference is involved in either case.

Finally in this section we consider a study on the role of verbalization in probability learning. Stevenson and Weir (1963) found that subjects responding in pairs did not perform differently from a single subject. Subjects who were forced to verbalize the basis of their responses did not differ in their choice behavior from those who did not verbalize explicitly. The sample in their experiment was comprised of 78 twelve-, fifteen-, and eighteen-year-olds and it is not certain that these findings can be assumed applicable at a younger age.

### III. CLASSROOM STUDIES OF MATERIALS AND INSTRUCTIONAL METHODS

Many studies have investigated how the ability to think in probability terms can be developed most effectively. Ojemann, Maxey and Snider (1965, 1966) devised a program of guided experiences designed to help the child learn elementary aspects of probability. A series of five 30-minute lessons were administered on consecutive days to a third grade class of 20 pupils with a corresponding class of 21 as the control group. One of the main concerns in designing materials and experiences for the lessons was that responses be encouraged which relate to the information available rather than only to previous experiences of interaction with the environment. Four tests were used to assess the effect of the instruction. Taken



together the results indicated that the experimental subjects were acquiring considerable ability to relate their "predictions" to the information available.

They showed significantly greater ability to relate their predictions to the probable and they tended to wait before making a prediction when only a small amount of information was available and more would be supplied. (p. 326)

The investigation seemed to indicate that grade three children can benefit from instruction in concepts of risk, chance maximization, and prediction.

At about the same time, and also in Iowa, Wilkinson and Nelson (1966) conducted a three week trial-teaching experiment with a sixth-grade class in the area of probability. The study sought answers to questions about content suitability and organization of activities. The class was taught for 45 minutes each day and this sequence appeared to be the first formal experience with probability encountered by any of the subjects. The most basic consideration was that experiences and ideas should be meaningful to the students. The lessons began at low levels of sophistication and activities were carried as far as possible while class interest was maintained.

The general technique was to initiate discussion with carefully stated questions and to encourage students to develop ideas through actual experimentation. Exploratory questions such as "Why?", "What does 'usually' mean?", "Are you sure?", or "How can you find out?" were employed and instructor responses were mainly neutral. Experience with eighteen concepts and five skills was incorporated into the



sequence of lessons which employed three probability situations: personalistic, familiar, and unfamiliar. The first involved data from the student's own environment such as birthdays, telephone numbers, and personal statistics. The second situation used coins, dice, and cards and the third dealt with objects such as thumb tacks, bent paper clips, paper cups, cardboard cylinders, and drawing beads from a container.

Approximately equal time was spent on each situation and questions were asked about likelihood, fairness, number of possible outcomes and combinations, and certainty. No concerted effort was made for precision in defining concepts; rather they were mostly dealt with on an intuitive level only.

In many of the experiences subjects realized the uncertainty of the situation and the need for a consensus of methods as indicated by a comment: "We've just got to make some agreements and then we can get somewhere" (p. 102). In dealing with familiar situations, the investigators were surprised at the number of inaccurate beliefs and intuitions held by the subjects concerning coins in particular.

When testing quarters to see if engraving differences cause tails to occur more often than heads, students were willing to make generalizations from ten flips if the results agreed with their prejudice. They said that six tails and four heads in ten trials would prove their point . . . More than six tails . . . was especially satisfying to them. Five tails . . . was treated as something you have to expect once in a while, but which doesn't disprove anything . . . Fewer than four tails in ten trials . . . led some to say that something had been dishonest (pp 102-103).

Less prejudice was found with dice and none with cards except that a few students were aware of the existence of trick



decks. No rationale was given by students for their objection to the use of a large die and a small die together but it seemed "unfair" to them. No such prejudice seemed to be held with regard to the unfamiliar situations. The students were encouraged in hypothesis-making and -testing activities which all agreed were interesting and worthwhile.

The important aspect of this experiment is the progression from a beginning estimate of a probability without necessarily being "exact" at any time (p. 105).

This would seem to be the essence of practical probability experiences: many questions in everyday life have answers of an indefinite character and a good guess is often the best answer possible.

Wilkinson and Nelson concluded with six recommendations that have implications for those involved in designing probability units for elementary school children. In brief summary these are:

1. Don't let intuition lead you too far.
2. Avoid pre-prejudiced situations.
3. Keep vocabulary useful and simple.
4. Spread the teaching out. Two- or three-day units spaced throughout the year would be more suitable than a three-week unit.
5. Don't overstructure the experiences.
6. Use experiences meaningful to students, experiences planned especially for them, not watered-down versions of high school or college-level probability.

Another significant study of teaching probability to sixth-





grade children was reported by Shepler (1969). He designed a four-week unit which was taught to a class of 25 students of average ability by a trained elementary school teacher. The instructional goal was to demonstrate "mastery learning" of the behavioral objectives of the unit.

Pre- and posttests were administered, the first to measure preknowledge of 14 objectives, the second to help both in re-analyzing the unit regarding modifications to be made and in testing the feasibility of the study. The test consisted of 72 items based on one- and two-dimensional finite sample spaces generated by models using coins, dice, spinners, and boxes of objects.

The criterion for instructional success was that 90% of the students should score 90% or better on each of the measured objectives. This was satisfied in the posttest in the case of 11 of the 14 objectives. There was a dramatic change in the performances on pre- and posttests with the mean scores increasing from 38% to 93%, and the variances decreasing from 74 to 11. The instruction was judged highly successful and the author attributed the large gain in raw score to the developmental analysis used and the mastery learning techniques employed.

The author proposed the following sequence for developing research-based curriculum materials:

Start with a content outline and establish behavioral objectives. Task analyse these objectives and write an instructional treatment to meet them. Proceed to the important step of actually trying these materials with children, while recognizing the possibility of iteration through preceding steps (pp. 202-203).



Shepler followed his own advice in analyzing the items which measured the three missed objectives. He concluded that poor wording in some items and lack of emphasis of objectives in others were plausible reasons for not achieving the objectives.

Lovell (1971), in commenting on the above study, stated "there is no doubt . . . that, given first-class teaching, selected sixth-grade pupils can be introduced to notions of probability" (p. 135). Despite its undoubted success, Shepler himself admitted the study had only narrow implications and was not generalizable. Based on his experience in training the teacher for the study he suggested that the typical elementary school teacher could adequately teach lessons on probability using a one-dimensional sample space. He cautioned, however, that more research and development needs to be done to determine more feasible ways of presenting problems in a two-dimensional sample space. As mentioned earlier in this report, reservations were also placed on the use of graphing situations calling for subtle interpretations.

A follow-up study of retention of probability concepts was conducted by Romberg and Shepler (1973) to examine the effects of the mastery learning technique utilized by Shepler in the study just reported. The same 72-item test was administered exactly four weeks after the posttest with no instruction or practice being given in that period. Posttest and retention test scores had a correlation of 0.78. The authors claimed a high retention rate. On the posttest 21 out of 25 (84%) had



achieved a 90% score. Seventeen of these 21 were still above a 90% level of performance and the other four were still above the 80% level.

The results also indicated that if the objective was originally mastered it was retained. If not, there was some loss. The eleven objectives successfully mastered in the instruction period had retention ratios in excess of 0.80; the other three objectives had ratios of 0.74, 0.54, and 0.43.

The authors acknowledged a weakness in such a study. The same test being used three times in a period of eight weeks could have caused test-retest interaction which was not controlled for. Also four weeks may not have been long enough to be practically significant. If further retention studies using parallel tests over longer periods were to corroborate the findings of their study, the investigators would then recommend further use of mastery-learning principles with this age group.

Two researchers did venture a little lower into the elementary grades in investigating the teaching of probability concepts. McLeod (1971) investigated the feasibility of teaching an eight- to ten-day unit to second- and fourth-grade children. He conducted two consecutive parallel studies. Experience gained in Study A was used to make modifications in instructional treatments and in the outcome measures in Study B. At both grade levels in a single school in each study, three experimental treatments were assigned at random to whole classes: laboratory participation (LP), teacher demonstration (TD), and no instruction





(M) between pretests and posttests. In study B control classes (C) were added from outside the basic school. Only the posttests were administered to these groups under C.

All the activities carried out in LP or observed in TD involved repeated chance events using the drawing of red and blue marbles from bags. The grade four activities were a little more extensive than those in grade two, otherwise the treatments were the same and common worksheets were used. Parallel 44-item tests were administered as pretests, posttests, and retention tests five weeks later.

Evidence from the pretests indicated that most second-grade as well as most fourth-grade children were able to apply the concepts of likely, more likely, equally likely, less likely, and unlikely before instruction began. Groups LP and TD were significantly superior to group C at both grade levels in Study B on the early posttest measures. No clear treatment effect was found for groups M, LP, and TD at either grade level on either posttest or retention measures. Learning apparently occurred under all three treatments. No clear effect was found for high or low groups classified on reading ability using Stanford Achievement Test scores and no effect was found due to sex.

The treatments LP and TD apparently improved the subjects' performance but so did M, no instruction apart from the pretest. This may have been due to a Hawthorn effect except that M was not significantly more effective than C, posttest only. Where knowledge is being measured rather than rate of production,



such a result tends to indicate that the knowledge is already present at the beginning. The pretest may have acted as an organizer of schema and ideas and consequent posttest performance was sufficiently improved so as to indicate no significant difference in the three treatment measures.

Rather than having strong implications about instructional procedures McLeod's study indicated the presence of a considerable range of understanding about chance situations in grade two and four children. Gipson (1971), at the same time, was examining just two concepts of probability and sought to give an accurate account of eight children's responses in learning the two concepts, finite sample space and probability of a simple event.

A procedure very similar to that outlined by Shepler (1969), reported earlier in this chapter was used by Gipson to develop an instructional sequence. Two pilot studies were used, the first to identify materials and appropriate concepts. In the second pilot study four children were taught on an individual basis and as a result three lessons were sequenced for presentation to third- and sixth-grade children. The lessons were taught to six children and audiotaped, and two more children were videotaped while being taught.

Pretest and posttest results and analysis of the children's protocols indicated the instructional sequence to be successful. The researcher reported that the interview-type procedure gave a deeper insight into how children think about probability concepts. The children were often able to explain how they



arrived at their answers relating to the two concepts. The conclusion was made that third grade is therefore an appropriate grade to introduce selected probability concepts.

Analysis of the protocols indicated that the most difficult performance objectives for the study were those related to specifying the estimated probability of the results for experiments and comparing equally likely outcomes using different objects. Gipson called for similar investigations to be made of children in grades four, five, and six.

It appears to the present investigator that Gipson's conclusion that third grade is the appropriate starting place for probability topics could be hasty. It would seem necessary to examine the situation with grades one and two more thoroughly before such a decision could properly be made. As indicated by studies reviewed in part II of this chapter, there is ample evidence that young children do have some understanding of some probability concepts. The nature and level of this understanding needs to be determined as a first step in designing curriculum material and instructional techniques suitable for the early grades.

Writers such as Harvey (1972) and Engel (1966, 1970) have provided detailed outlines of topics within probability that should be taught at upper elementary and junior high school levels. Engel's approach is a set theoretical one and rapidly becomes too advanced for most elementary students. Harvey, on the other hand, begins with descriptive statistics and generates behavioral objectives for over 130 tasks. These also



seem to be pitched at upper elementary grades and are very statistics-oriented, but some of the earlier tasks appear suited to junior grades.

#### IV. SUMMARY

Piaget's experiments led him to propose three stages in the development of probabilistic thinking in children. In the first stage, four to seven years, children have little or no idea of random mixtures and fortuitous events. They tend to impute a hidden lawfulness to random processes and to explain outcomes in terms of egocentric and quasi-magical causal relations. In the second stage, seven to eleven years, children begin to recognize randomization and irreversibility in probabilistic situations and can understand most probability concepts. Piaget maintains that not until the third stage, beginning at age eleven or twelve, can the child really understand the process of random mixture and deal with the quantification of probability.

The main criticism of Piaget's theory relates to his first-stage proposal. Several studies have shown that young children, even of preschool age, can learn basic ideas about probability if the instructional conditions are reinforcing and the experiences are meaningful to the pupils. Other research has found that school children of all ages appear to develop concepts of probability prior to receiving any formal instruction in the topic. The intuitions about probability that children have are often incorrect and need to be determined and understood before





a meaningful program can be designed.

Classroom studies of material and instructional methods have proven the feasibility of teaching probability topics in elementary grades. At the same time these studies have shown the value in adopting a behavioral-objectives and mastery-learning approach to developing instructional units.

There is general agreement with Piaget that quantification of probability seems unlikely until age eleven or twelve although Fischbein reported that nine-year-olds had been successfully taught combinatoric skills which enabled them to compare ratios in probability situations.

Most studies showed that age and instructional conditions were significant factors in children's performance on probability items. Socioeconomic status and IQ were often found to be significant though no strong generalization could be made from all studies collectively.

Only Jones (1975) seems to have investigated the level of probability concepts at the grade one, two, and three levels. This present study was in part a replication of his study in that the effect of embodiments was investigated with these same grades and a game situation was used as motivation to maximize strategies. Choice between arrays was the required response for one part of the investigation and prediction of the estimated probability was required in another part. While the test required largely non-verbal response, found by Yost and others to be preferable with young subjects, verbalization of the basis for response was encouraged throughout the interviews.



Every attempt was made to fit the task-conditions to the subject's level of development in order to maximize their responses.

Details of the design are given in the next chapter.



## CHAPTER 3

### INSTRUMENTATION AND RESEARCH PROCEDURES

#### I. INSTRUMENTATION

To test the hypotheses stated in Chapter 1 it was necessary to construct and administer an appropriate probability test and to administer the Canadian Cognitive Abilities Test. These instruments are described below.

##### Probability Test

Purposes. The purposes of the probability test were:

1. to obtain a measure of each subject's understanding of the six probability concepts in question,
2. to explore the extent to which subjects were able to quantify the probabilities in the settings presented to them, and
3. to determine the kinds of reasons that subjects gave for their responses.

Each concept was presented in three embodiments using spinners, blocks, and boxes. Throughout the test a total of five different probability settings were employed to elicit responses about the concepts being examined. Three of the settings were used in examining subjects' ability to quantify a probability representation.

Materials and apparatus. Three sets of devices were made:

1. five spinners each with equal sectors outlined in black





and colored with plastic adhesive tape,

2. five 2 cm wood blocks covered with the same plastic adhesive tape, and

3. five plastic boxes each containing six plastic counters, which were the same shades of color as the adhesive tape.

The proportions of blue, red, and yellow for each of the fifteen devices in the test are given in table 1.

TABLE 1

PROPORTIONS OF BLUE, RED, YELLOW IN THE DEVICES  
USED IN THE PROBABILITY TEST

| Device  | Proportions (B:R:Y) |       |       |       |       |
|---------|---------------------|-------|-------|-------|-------|
| Spinner | 2:2:2               | 3:2:1 | 4:1:1 | 3:3:0 | 0:6:0 |
| Block   | 2:2:2               | 2:3:1 | 1:4:1 | 3:0:3 | 6:0:0 |
| Box     | 2:2:2               | 1:2:3 | 1:1:4 | 0:3:3 | 0:0:6 |

A white laminated race-game board was made (see Appendix A) and a collection of six markers was used, one of each color blue, red, yellow, white, green, and brown. The researcher was careful to arrange the materials so that counters, markers, and plastic tape were the same shade of blue, red, and yellow.

For use in the introductory activity, a half green half white spinner and a half red half yellow block were also made. Except for these two, the devices were assembled so that no one color appeared to be used more often than another and color, devices, and correct response in the test were randomly associated.



Responses. Three types of response were sought from each subject corresponding to the three purposes of the test.

1. Choice response, to measure understanding of the six concepts. For example, "This spinner" (pointing to it) or "This marker" (picking it up or touching it). Twenty one such responses comprised the concept subtest.

2. Predictive response, to measure subjects' quantitative understanding of probability. For example, "Four times out of six goes I'll get yellow" although "Four" was the usual abbreviated version of such a response. Eighteen such responses sought from each subject comprised the quantitative subtest.

3. Rationalization response, in answer to the question "Why did you choose that one?" after each type (1) response, or "Why that number of times?" after each type (2) response. Three possibilities existed at this point:

(a) A rational, correct explanation was given based on the probability settings present, for example "it has more blue than red or yellow" or "there are four red sides on this block so we will be likely to get red four times out of six".

(b) A non-rational explanation was given involving influences such as favoritism of color, position, or quasi-magical properties attributed to a color or to a device.

(c) No response was given or the child said "I don't know" or simply "No reason".

A subject was never pressed for an answer beyond one repetition of the question and every effort was made to ensure that the subject did not feel threatened but in fact enjoyed the



interview. The researcher was encouraged to believe he succeeded in this by the number who expressed a wish to return and "play again" and by the number of children not in the sample who approached him at the school to request inclusion.

### Description of the Probability Test

Introductory activity: the game. As most of the questions in the test were presented in the setting of a simple race-game the interview began with an introduction to each other and to the game. The researcher placed the game-board, the green/white spinner, the green marker, and the white marker in front of the subject who was asked if he liked to play games. All students answered in the affirmative. Most had seen a spinner before but only the type with numerals on the face. The researcher explained and demonstrated how the two markers were used, that one advanced one square on the board when the spinner stopped on its color. The first to reach "finish" was the winner. The subject was invited to choose one of the markers and the game was played to a conclusion with either subject or investigator operating the spinner.

The spinner and markers were removed and the red/yellow block and matching markers were placed on the board. After a brief examination of the block and its use in the game (e.g. red on the uppermost face means the red marker moves up one square) the subject and investigator played the game again to a conclusion. By that stage all subjects said they knew how to play the game and the test was then begun.



The test organization and design. The test contained fifteen items which were related to the concepts, quantification trials, and embodiments as shown in Table 2.

TABLE 2

ITEMS AND EMBODIMENTS BY WHICH CONCEPTS  
AND QUANTIFICATION WERE TESTED

| Concepts                       | Items                              | Embodiment*             |
|--------------------------------|------------------------------------|-------------------------|
| Sample space events            | 1 (a), (b); 2 (a), (b); 3 (a), (b) | 2-2-2                   |
| Most favorable event           | 4 (a), 5 (a), 6 (a)                | 3-2-1                   |
| Most favorable sample space    | 7 (a), 8 (a), 9 (a)                | 4-1-1                   |
| Equally favorable sample space | 10 (a), 11 (a), 12 (a)             | 2-2-2                   |
| Impossible event               | 13 (a), 14 (a), 15 (a)             | 3-3-0                   |
| Certain event                  | 13 (b), 14 (b), 15 (b)             | 6-0-0                   |
| Quantification                 |                                    |                         |
| Six trials                     | Part (b) of 4 through 12           | 3-2-1<br>4-1-1<br>2-2-2 |
| Twelve trials                  | Part (c) of 4 through 12           | 3-2-1<br>4-1-1<br>2-2-2 |

\* each entry represents the three settings which are isomorphic to the given one (e.g. 3-2-1 means 3-2-1, 2-3-1, and 1-2-3 were used).





Items one, two, and three were presented, in random order, followed by items four through fifteen, again in random order. In the first three items more effort was made to "draw out" the desired response than in later items. It was felt that the concept of sample space was basic to all the others. Without an appreciation of what the possible events in a given situation were, a subject would find later items unnecessarily difficult if not meaningless. Subjects were not asked for rationalizations for their responses on these first three items.

From Table 2 it can be seen that a total of 21 questions was asked relating to concepts and 18 relating to quantification predictions. The items are now described in detail.

#### Description of the Test Items

##### Item 1.

(a) The researcher placed the six markers near the game board. The 2-2-2 spinner was shown to the subject who was asked "If this spinner were used to play the game which of these markers would be used?" Usually the subject selected the three correct markers and placed them in the starting squares on the board. If only one marker was selected, the question "Is that all?" was asked until the subject gave a firm "yes".

(b) With just the 2-2-2 spinner before the subject, the researcher asked "If I were to spin this spinner what could I get?" With few exceptions subjects responded correctly in one of two forms. For example, "A red, or a blue, or a yellow" or "You could get any of the colors". The probing question "Anything else?" was



allowed once with each subject if needed.

Item 2.

(a) The board and six markers were presented as described in Item 1(a) but this time the 2-2-2 block was used and the question adjusted to read "block" instead of "spinner".

(b) With just the 2-2-2 block before the subject the researcher asked, "If I were to roll this block what could I get?" The same procedure for probing as in item 1 was employed as needed.

Item 3.

(a) The game board and six markers were presented as in items 1(a) and 2(a). The 2-2-2 box was presented and its use explained to the subject. When the researcher felt sure the subject understood how the box could be used to play the game he asked the question, "If this box were used to play the game which of these markers would be used?".

(b) With just the 2-2-2 box before the subject the researcher asked "If I were to reach into this box and draw out one counter without looking what could I get?". The same probing procedure as described in item 1 was employed as needed.

In each of the following items the game board was before the subject with a blue, a red and a yellow marker in the start squares. Each question relating to the game was in the context of these three "players".

Item 4.

(a) The 3-2-1 spinner was presented and the subject was asked "If using this spinner which marker would you choose to



win the game?". When a response was given the subject was then asked "Why that one?".

(b) The subject was then asked "In six spins how many times would you expect to get blue?". If a response was given the subject was encouraged to state a reason for his response by the question "Why (the number)?".

(c) The subject was then asked "In twelve spins how many times would you expect to get blue?". Again, if he responded, his reason was elicited by the question "Why (the number)?".

#### Item 5.

This item was identical to item 4 except that the embodiment was the 2-3-1 block and the wording of the questions was altered accordingly:

(a) "If using this block which marker would you choose to win the game?"

(b) "In six rolls how many times would you expect to get red?"

(c) "In twelve rolls how many times would you expect to get red?"

#### Item 6.

This item was identical to item 4 except the embodiment was the 1-2-3 box and yellow was the color most favored. The questions were worded as follows:

(a) "If using this box which marker would you choose to win the game?"





(b) "In six draws how many times would you expect to get yellow?"

(c) "In twelve draws how many times would you expect to get yellow?"

Item 7.

(a) The 2-2-2, 3-2-1, and 4-1-1 spinners were placed before the subject who was asked "Which spinner gives blue the best chance of winning?". When a response had been given, the question was asked "Why that one?".

(b) Referring only to the 4-1-1 spinner the researcher asked "In six spins how many times would you expect to get blue?". The 2-2-2 and 3-2-1 spinners were removed from view prior to putting this question in order to encourage the subject to attend only to the 4-1-1 setting. The subject's reason for any response was sought by asking "Why (the number)?".

(c) The subject was then asked with regard to the 4-1-1 spinner only "In twelve spins how many times would you expect to get blue?". Again the follow up probe "Why (the number)?" was used.

Item 8.

This was a replication of item 7 in the block embodiment.

(a) The 2-2-2, 2-3-1, and 1-4-1 blocks were presented and the subject was asked "Which block gives red the best chance of winning?" and "Why that one?".

(b) Only the 1-4-1 block was left before the subject who was asked "In six rolls how many times would you expect to



get red?". Upon giving a response the subject was asked "Why (the number)?".

(c) Referring only to the 1-4-1 block the researcher asked "In twelve rolls of this block how many times would you expect to get red?" and "Why (the number) times?".

Item 9.

This was the box embodiment of item 7.

(a) The 2-2-2, 1-2-3, and 1-1-4 boxes were presented and the subject was asked "Which box gives yellow the best chance of winning?".

(b) Referring only to drawing from the 1-1-4 box, the researcher asked "In six goes how many times would you expect to get yellow?".

(c) Again referring to box 1-1-4 the question was asked "In twelve goes how many times would you expect to get yellow?". Each time a response was given the subject's reason for that response was elicited by the question "Why that one?" or "Why that number?".

Item 10.

(a) The 2-2-2, 3-2-1, and 4-1-1 spinners were presented to the subject who was asked "Which spinner gives each player (indicating the three markers in place on the game board) the same chance of winning?". Occasionally a subject appeared not to understand the question whereupon the researcher rephrased the question to "Which spinner makes the game fair for each of the three players?". The subject's reason for his response was



again sought by the question "Why this one?".

(b) The 3-2-1 and 4-1-1 spinners were removed from the subject's view and, referring only to the 2-2-2 spinner, the subject was asked "In six spins how many times would you expect to get red?" and "Why that number?".

(c) Again referring only to the 2-2-2 spinner the subject was asked "In twelve spins how many times would you expect to get red?" and "Why that number?".

#### Item 11.

The procedure was as described in item 10 except that the 2-2-2, 2-3-1, and 1-4-1 blocks were used in (a), yellow was the chosen color in (b) and (c), and appropriate wording was used in the questions (e.g. "rolls" instead of "spins"). The subject was always given the opportunity to state his reason for his response as described in previous items.

#### Item 12.

The procedure was as described in item 10 except that the 2-2-2, 1-2-3, and 1-1-4 boxes were used in (a), blue was the chosen color in (b) and (c), and appropriate changes were made to the terms in the questions (e.g. "draws" or "goes" instead of "spins"). Again the probe questions "Why that one?" or "Why (the number)?" were always asked.

#### Item 13.

(a) The 2-2-2, 3-2-1, 4-1-1, 3-3-0, and 0-6-0 spinners were placed before the subject who was instructed to look



carefully at all of them. The researcher said "We have these three players (pointing to the markers) in the game. Which one of these spinners makes it so that just one of the players can never win?". This was repeated carefully with the alternative "will always lose" being given at the end. When a spinner was selected the subject was asked "And which marker will always lose if we use this spinner?". Following a nomination of marker(s) the researcher asked "Why?".

(b) Restoring the spinner selected in (a) to the array of five spinners the researcher said, "Look carefully at the spinners and tell me which spinner makes it so that one of these players will always win?". This was repeated if necessary and again the subject was required to indicate a spinner and a marker and to give a reason for the choice.

Item 14.

The 2-2-2, 2-3-1, 1-4-1, 3-0-3, and 6-0-0 blocks were placed before the subject. The questions (a) and (b), with probes, were asked as described in item 13 with the appropriate change of wording to "blocks" instead of "spinners".

Item 15.

The 2-2-2, 1-2-3, 1-1-4, 0-3-3, and 0-0-6 boxes were placed before the subject. Questions (a) and (b) and probes as described in item 13 were asked using the word "boxes" instead of "spinners".





### Pilot Study of the Probability Test

A draft form of the test was piloted on March 23, 1978 in a single Edmonton school using three subjects from each of grades one, two, and three. The test was administered individually to each subject in a manner similar to that described in the preceding paragraphs. A seven inch reel-to-reel tape recorder was used to record the interviews. The first draft of the test contained only items which investigated the six concepts in question but not quantification.

The purpose of the pilot study was to trial the items, the materials, the interview procedure, and the recording technique. In addition it helped to satisfy the researcher that the instrument was appropriate to the grade levels involved and that the items were valid measures of the elementary probability concepts. Feedback from a panel of six independent mathematics educators to whom the findings of the pilot study were reported supported the claim for validity of the items.

Two major changes were made as a result of the pilot study. The tape recorder provided marginal assistance in the collecting of data and was therefore not used in the main study. The responses given by the subjects were easily categorized, coded, and recorded by the researcher on an answer sheet and the tape added little additional information. Secondly it was decided to inquire further into the subjects' probabilistic understanding by including the quantitative parts of the test items which were not included in the pilot form of the test (parts (b) and (c) of items 4 through 12). These were then trialled separately on the



researcher's own three children, ages nine, seven, and five years, mainly to streamline the actual wording of the questions and the presentation of the materials.

The pilot study proved to be of great value also in providing the investigator with practice in interviewing and recording with young children.

#### Canadian Cognitive Abilities Test (CCAT)

The second instrument used in the study was for the purpose of gaining a measure of the subject's general reasoning ability. As grade three subjects in the school system had recently been tested with the Canadian Cognitive Abilities Test it was decided to administer the companion tests in the same series to the grade one and two subjects.

"The Cognitive Abilities Test is part of an integrated test series designed to assess the cognitive development of abilities from kindergarten through grade nine" (Thorndike, 1968, p. 4). Primary 1 and Primary 2, designed for grades one and two respectively, are group tests using pictorial materials and oral instructions. There are four short subtests in each, oral vocabulary, relational concepts, multi-mental ("one that doesn't belong"), and quantitative concepts.

Norms were established for CCAT Primary in 1966 using 50 schools across Canada supplying approximately 2 000 pupils in each of grades K to 4. Construct validity was established using a factor analysis which showed that all batteries in the series give measures of a general reasoning factor. Reliability



coefficients of from 0.769 to 0.887 are given for the tests.

A table is supplied to allow easy conversion of raw scores and chronological age into deviation IQ's with a mean of 100 and standard deviation of 16. These are standard scores directly comparable from age to age.

## II RESEARCH PROCEDURES

### Selection of Sample

The Edmonton Public School Board made one school available for data collection. It was situated in a rapidly developing residential neighborhood regarded as representative of Edmonton's middle to low socio-economic status areas. The principal of the school was contacted and asked to make 24 pupils available to the researcher from each of grades one, two, and three. A further request was made for the sample to have equal numbers of boys and girls and equal numbers of pupils of high and low general reasoning ability.

As no standardized IQ scores were available for grade one and two pupils the principal suggested that the teachers' perception be the means of judging the general ability levels of all of the subjects. In this way the sample of 72 subjects was selected from a population of 176 grade one, two, and three pupils. The principal had the list of names of subjects ready for the researcher on his arrival at the school for data collection.





### Data Collection

The data were collected in the last two weeks of April, 1978. The researcher, assisted by his wife, spent eight days at the school interviewing subjects individually and administering the Cognitive Abilities Test in several sittings to the grade one and two subjects.

The probability test was administered on an individual basis in a small room made available by the principal. The procedure was as follows: familiarize the subject with the game by playing it as outlined previously, present the items 1, 2, and 3 in a random order, and present items 4 through 15 in a random order.

To facilitate this procedure question cards were made for each of the fifteen items. Each card indicated the item number, listed the apparatus to be used and showed each part of the question as it was to be stated. Labelling of the apparatus with the appropriate tri-numeral description of the probability setting enabled the researcher to quickly change the materials between items.

After each interview the researcher or assistant ordered the cards according to the next row in a table of random numbers generated especially for the purpose. In this way an attempt was made to control for a teaching effect which might otherwise influence scores in items occurring towards the end of the test.

The average time for an interview was approximately twenty minutes. Twelve subjects were tested each day, usually seven



in the morning and five after lunch.

Each subject's responses were recorded by the researcher on individual profiles. These responses were later coded, along with all the identification information, onto a data form using a one or zero according to whether the response was correct or not. The rationalizations for responses given throughout the interviews were noted and later were tallied into a frequency table for analysis. The range of replies was sufficiently narrow for the researcher to classify them into three groups as described in the first section of this chapter.

The remaining data, age and IQ were secured from the cumulative record cards in the case of grade three subjects. For grade one and two subjects, as mentioned above, the IQ had to be obtained from direct testing by the researcher. As the school would have had little use for IQs for grade one and two pupils at that late stage in the school year the principal preferred that only the sample subjects be tested.

The Primary 1 Form 1 test in the Canadian Cognitive Abilities Test series was administered to the 24 grade one subjects by the researcher and his assistant in three thirty minute sittings over two days. The Primary 2 Form 1 test in the same series was similarly administered to the 24 grade two subjects. The IQs were derived according to the test manual and were added to the subject's data card.

Finally the occupation of each subject's father was noted from the cumulative record cards for the purpose of gaining an average occupational class rating of the sample (Blishen, Jones,



Naegele, & Porter, 1968). The Blishen index was based on income and education characteristics of incumbents of 320 occupations in the 1961 Canadian census and is utilized in this study simply as a check on the judgment of both the researcher and the school principal as to the socio-economic status of the sample. An average index of 43 on a scale ranging from 25 to 77 confirms the assessment of middle to low SES which was stated earlier.

### Analysis of Data

The data were analysed with the assistance of the computer facilities in the Division of Educational Research Services at the University of Alberta. The encoded data were entered into a file and several standard statistical programs were utilized.

Test statistics were calculated initially for the total sample and these are reported in the analysis of the instrument below. Further analysis of the test responses included item frequencies, responses according to embodiment, and means and variances within IQ, grade, and sex groups. One- and three-way analyses of variance were performed to test for effect of embodiment and effect due to sex, grade, and IQ.

These analyses are reported and discussed in the next chapter along with an analysis of subjects' rationalizations.

### III. ANALYSIS OF THE PROBABILITY INSTRUMENT

This section reports a post-administration analysis of the probability test using the data collected in the present study.

Table 3 contains the means and standard deviations for the



total test scores and two subtest scores for the whole sample and for each grade.

TABLE 3  
MEAN AND STANDARD DEVIATION ON PROBABILITY TEST AND  
SUBTESTS USING TOTAL SAMPLE AND GRADES

|              | Total score |      | Concept   |      | Quantification |      |
|--------------|-------------|------|-----------|------|----------------|------|
|              | $\bar{X}$   | S.D. | $\bar{X}$ | S.D. | $\bar{X}$      | S.D. |
| Total sample | 22.81       | 5.91 | 17.21     | 3.31 | 5.63           | 4.10 |
| Grade 1      | 19.88       | 6.09 | 15.13     | 3.44 | 4.83           | 4.37 |
| Grade 2      | 21.79       | 4.66 | 17.08     | 3.15 | 4.71           | 3.42 |
| Grade 3      | 26.75       | 4.57 | 19.42     | 1.35 | 7.33           | 3.91 |

The three scores are out of 39, 21, and 18 respectively and the means indicate the relative difficulty of the quantitative items.

An analysis of variance indicated that grade was a significant factor in all three criterion measures. The Scheffe method of multiple comparisons of means was used to compare the performance of grade levels two at a time (Ferguson, 1971, p. 270-271). The probabilities derived from the Scheffe test are given in Table 4. Significant differences at the 0.01 level occurred between the total score means for grades one and three and grades two and three. The grade one and grade three means were also significantly different at the 0.01 level on the concept subtest.





TABLE 4

PROBABILITIES FOR SCHEFFE MULTIPLE COMPARISONS OF  
GRADE MEANS ON TOTAL AND SUBTEST SCORES

| Grade means<br>compared | Test   |         |                |
|-------------------------|--------|---------|----------------|
|                         | Total  | Concept | Quantification |
| 1 vs 2                  | 0.4553 | 0.0675  | 0.9942         |
| 1 vs 3                  | 0.0001 | 0.0000  | 0.1039         |
| 2 vs 3                  | 0.0071 | 0.0231  | 0.0831         |

The distribution of scores on the test is shown in Figure 1. For the total sample the distribution of scores was found to be normal by the chi-square goodness of fit test. The skewness was -0.028 and the kurtosis was -0.268. The chi-square of 5.744 with 6 degrees of freedom had a probability of 0.453. Neither of the subtests was indicated to be normally distributed, the chi-square probabilities being less than 0.001 and 0.006 for concept and quantification subtests respectively. These measures are summarized in Table 5.

TABLE 5

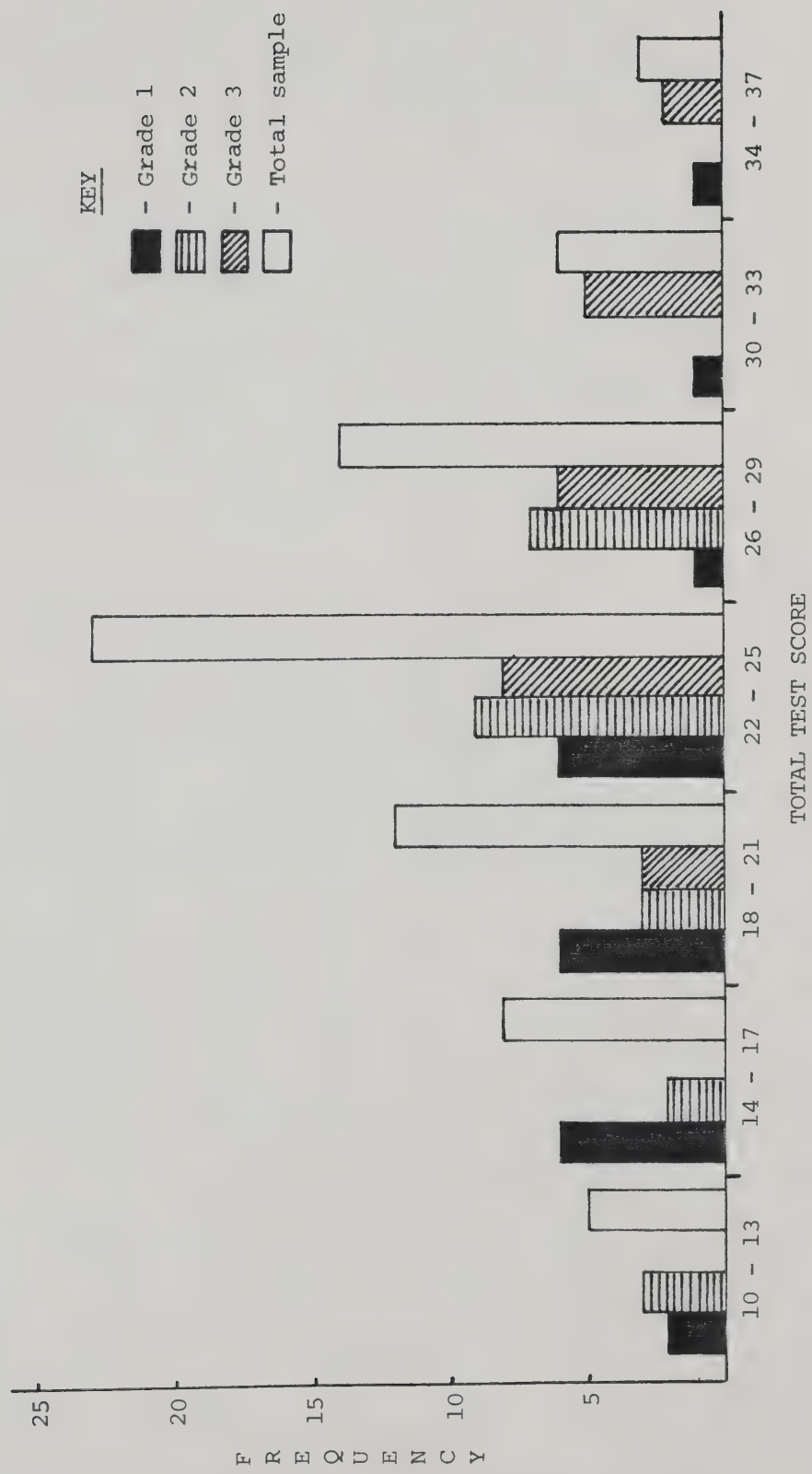
SKEWNESS, KURTOSIS, AND CHI-SQUARE  
FOR TEST AND SUBTESTS ON TOTAL SAMPLE

| Test           | Skewness | Kurtosis | $\chi^2$ | df | Prob. of $\chi^2$ |
|----------------|----------|----------|----------|----|-------------------|
| Total          | -0.028   | -0.268   | 5.744    | 6  | 0.453             |
| Concept        | -1.069   | 0.858    | 19.699   | 4  | 0.001             |
| Quantification | 0.722    | 0.182    | 10.146   | 2  | 0.006             |



FIGURE 1

DISTRIBUTION OF SCORES ON THE PROBABILITY TEST BY GRADES AND TOTAL SAMPLE





As a measure of item difficulty the percent of responses correct was calculated for each item. These percentages are presented in Table 6 where the items are grouped in the two subtests.

The concept items were correctly answered at least 73% of the time except for those relating to concept five, the impossible event. By contrast the percentages for the quantification items were less than 50% except for item 9(b). More discussion of particular groupings occurs in chapter 4.

TABLE 6

## PERCENTAGES CORRECT ON PROBABILITY

## TEST ITEMS FOR TOTAL SAMPLE

| Concept items |     |      |    | Quantification items |    |      |    |
|---------------|-----|------|----|----------------------|----|------|----|
| Item          | %   | Item | %  | Item                 | %  | Item | %  |
| 1 a           | 100 | 9 a  | 96 | 4 b                  | 41 | 9 b  | 53 |
| b             | 99  | 10 a | 74 | c                    | 31 | c    | 17 |
| 2 a           | 98  | 11 a | 73 | 5 b                  | 42 | 10 b | 41 |
| b             | 98  | 12 a | 77 | c                    | 32 | c    | 12 |
| 3 a           | 96  | 13 a | 57 | 6 b                  | 45 | 11 b | 41 |
| b             | 99  | b    | 81 | c                    | 25 | c    | 16 |
| 4 a           | 92  | 14 a | 46 | 7 b                  | 35 | 12 b | 46 |
| 5 a           | 80  | b    | 87 | c                    | 13 | c    | 19 |
| 6 a           | 81  | 15 a | 46 | 8 b                  | 34 |      |    |
| 7 a           | 91  | b    | 85 | c                    | 13 |      |    |
| 8 a           | 82  |      |    |                      |    |      |    |





### Reliability

A test-retest reliability check was made on the final version of the probability test. Eighteen of the subjects (25% of the sample) were randomly chosen for retesting which was done in one day in the third week of May, 1978, three weeks after the original testing. All items were administered in the same interview situation as employed initially. Pearson product-moment coefficients of correlation were calculated on test and retest total, concept, and quantification scores. The correlation coefficients and test and retest means and standard deviations are given in Table 7.

TABLE 7

#### TEST-RETEST RELIABILITY OF PROBABILITY TEST

| Criterion score | Test      |      | Retest    |      | Correlation coefficient |
|-----------------|-----------|------|-----------|------|-------------------------|
|                 | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. |                         |
| Total           | 23.6      | 5.27 | 24.4      | 4.78 | 0.837                   |
| Concept         | 17.1      | 2.88 | 18.5      | 2.27 | 0.796                   |
| Quantification  | 6.6       | 3.67 | 5.9       | 3.02 | 0.753                   |

The differences between the test and retest parameters were judged to be small enough and the correlation large enough for the test to be considered reliable for experimental use.



## CHAPTER 4

## RESULTS OF THE INVESTIGATION

This chapter reports the results of testing the hypotheses and of other analyses made on the data. The results are reported separately for each of the four major purposes stated in Chapter 1. The questions and hypotheses listed under each purpose are discussed in turn.

## I. STATUS OF THE PROBABILITY CONCEPTS

Question One Restated: What proportion of subjects in each grade and in the total sample indicate an understanding of the six concepts investigated?

At the end of the previous chapter an indication of item difficulty was provided in Table 6 as the percentage correct on each item in the probability test for the whole sample of 72 subjects. To arrive at a status index for each probability concept embodied in the test items average relative frequencies were computed for the set of items relating to each of the six concepts. For concept one, items one, two, and three (two parts in each) constituted the set of related items. Each of the other five concepts was tested by three items, one for each embodiment. The relative frequencies for each concept are given as percentages in Table 8 for the whole sample and for each grade.

The percentages reported in Table 8 indicate that the first concept, sample space, was understood by almost all subjects,



whereas only half of the responses in the whole sample were correct to items about concept five, impossible event. The other concepts ranked between these extremes in the following order: most favorable sample space, certain event, most favorable event, and equally favorable sample space.

TABLE 8

MEAN PER CENT CORRECT RESPONSES FOR PROBABILITY CONCEPTS  
IN EACH GRADE AND THE WHOLE SAMPLE

| Concept                        | Grade (N=24) |    |    |     |    |     | Total<br>Sample<br>(N=72) |    |
|--------------------------------|--------------|----|----|-----|----|-----|---------------------------|----|
|                                | 1            |    | 2  |     | 3  |     |                           |    |
|                                | n            | %  | n  | %   | n  | %   | n                         | %  |
| Sample space                   | 23           | 96 | 24 | 100 | 24 | 100 | 71                        | 99 |
| Most favorable event           | 18           | 75 | 19 | 79  | 23 | 96  | 60                        | 83 |
| Most favorable sample space    | 18           | 75 | 23 | 96  | 23 | 96  | 64                        | 89 |
| Equally favorable sample space | 14           | 58 | 17 | 71  | 22 | 92  | 53                        | 74 |
| Impossible event               | 9*           | 38 | 11 | 46  | 16 | 67  | 36                        | 50 |
| Certain event                  | 18           | 75 | 20 | 83  | 23 | 96  | 61                        | 85 |

\* Not significantly better than chance (0.01)

The increase in percentages on all concepts through grades one to three indicates that these concepts develop with age.

Analysis within grades indicates that four of the concepts



were understood by at least 75% of the subjects irrespective of grade. These were concepts one, two, three, and six. If 75% were taken as a minimum criterion for acceptable level of probability concepts before formal instruction should begin, then this study indicates that four of the six concepts investigated meet that criterion in each grade and in the total sample.

Hypothesis Two Restated: The proportions derived in answering question one are not significantly different from chance proportions for each concept.

In order to determine the number of correct responses beyond the number obtainable by chance, the cumulative binomial distribution was used. The items relating to concepts two, three, and four had three possible outcomes and those relating to concepts five and six gave five alternatives. If a subject were to respond at random, the probability ( $p$ ) of a correct response based only on a random choice would be  $1/3$  and  $1/5$  respectively. With  $p=1/3$  and  $N=72$ , the binomial probability of any item obtaining more than 33 correct responses is less than 0.01. With  $p=1/5$  and  $N=72$ , the critical number is 23. Table 8 reveals that all the concepts received more than the critical number of correct responses from the whole sample.

The binomial probabilities were also found for the number in each grade in the sample. For  $p=1/3$  and  $N=24$ , the probability of any item obtaining more than 13 correct responses by chance is less than 0.01. For  $p=1/5$  and  $N=24$ , the critical number is 10





responses. Table 8 shows that only one concept has less than the significant number of correct responses in any of the grades. This is concept five, impossible event, at the grade one level. With this one exception, the hypothesis is not accepted. The conclusion is that all six concepts received correct responses in the whole sample significantly more often than is attributable to chance.

## II. LEVEL OF QUANTIFICATION OF PROBABILITY

Question Three Restated: What proportions of subjects in each grade and in the total sample indicate an understanding of the quantification items presented?

The data in Table 6 were used to compute the mean per cent of correct responses to the three questions on each of the six quantification situations. Table 9 presents the mean number and per cent of correct responses for each grade and for the total sample. As expected the items on quantification were much more difficult than those on the concepts. This is indicated by the lower percentages throughout Table 9, the highest rate of success being 58% achieved by grade three subjects in predicting the expected frequency of an event in six trials using the 2-2-2 settings.

Performance improved with the age of subjects except for a decline in grade two on two of the settings. The average percentages of correct responses indicate little difference between grades one and two with averages of 25% and 24% respectively,



but grade three subjects performed significantly better with an average of 40% over all situations. (A more complete analysis of differences between grades is deferred until hypothesis eight is discussed.)

TABLE 9

## MEAN PER CENT CORRECT RESPONSES ON QUANTIFICATION

## ITEMS IN EACH GRADE AND THE WHOLE SAMPLE

| Probability<br>setting | Grade (N=24) |    |    |    |    |    | Whole<br>Sample<br>(N=72) |    |
|------------------------|--------------|----|----|----|----|----|---------------------------|----|
|                        | 1            |    | 2  |    | 3  |    |                           |    |
|                        | n            | %  | n  | %  | n  | %  | n                         | %  |
| 2-2-2                  |              |    |    |    |    |    |                           |    |
| 6 trials               | 8*           | 33 | 8* | 33 | 14 | 58 | 30                        | 42 |
| 12 trials              | 3*           | 13 | 1* | 4  | 6* | 25 | 11*                       | 15 |
| 3-2-1                  |              |    |    |    |    |    |                           |    |
| 6 trials               | 8*           | 33 | 10 | 42 | 12 | 50 | 30                        | 42 |
| 12 trials              | 4*           | 17 | 6* | 25 | 11 | 46 | 21                        | 29 |
| 4-1-1                  |              |    |    |    |    |    |                           |    |
| 6 trials               | 9*           | 38 | 8* | 33 | 11 | 46 | 29                        | 40 |
| 12 trials              | 4*           | 17 | 2* | 8  | 4* | 17 | 10*                       | 14 |

\* Not significantly better than chance (0.01)

Hypothesis Four Restated: The proportions derived in question three do not differ significantly from chance proportions.

In answering the questions concerning quantification, the subjects were asked to predict how many times in six or twelve



trials a certain event would occur. If a subject were to respond in a completely random manner, the chance of his response being the correct one would be  $1/6$  or  $1/12$  respectively.

On this basis, the cumulative binomial distribution was used to determine the number of correct responses beyond the number obtainable by chance. With  $p=1/6$  and  $N=72$ , the binomial probability of any item obtaining more than 20 correct responses is less than 0.01. With  $p=1/12$  and  $N=72$ , the critical number is 12. Table 9 shows that, for the whole sample, responses on all three settings were significantly different from chance responses for the six-trial items but significant only on the 3-2-1 setting for the twelve-trial items.

Similarly, the critical points were found for  $N=24$  and  $p=1/6$  and  $1/12$ , so that within each grade the number of correct responses obtainable by chance alone could be compared to the performance of subjects in that grade. For the six-trial items the binomial probability of any item obtaining more than nine correct responses is less than 0.01, while the critical number for the twelve-trial items is six. Table 9 shows that 13 of the 18 mean responses within the three grades are not significantly better than chance at the 0.01 level of confidence.

For grade one, hypothesis four should be accepted for all settings and trial sizes. For grade two, the hypothesis should be accepted except for the six-trial 3-2-1 setting. For grade three the hypothesis is not accepted except for the twelve-trial questions with the 2-2-2 and 4-1-1 settings.

In view of the above analysis, it would be unwise to judge





this hypothesis solely on the total sample response as it is biased by the significantly higher correct response rate of the grade three students. This analysis supports the conclusions stated regarding question three that subjects in grades one and two performed poorly on the quantification items while over 50% of the grade three subjects indicated understanding in the six-trial situations.

An indication of the distribution of the actual responses to the quantification items is given in Tables 17 and 18 in Appendix B.

### III. THE EFFECT OF EMBODIMENT

Hypothesis Five Restated: There are no significant main effects due to embodiment, on performance on the probability test.

The hypothesis was tested using a one-way analysis of variance with repeated measures. Each concept and quantification question was essentially asked three times, once with each embodiment of isomorphic probability settings. The embodiments can thus be regarded as repeated measures of the same criterion.

The analysis of variance was carried out for each grade and for the total sample. The results of the comparison of mean scores on each embodiment, within grades and within the whole sample, are given in Table 10.

The conservative probability of F was calculated by making allowance for unequal covariances among the correlated measures (Winer, 1971, pp. 281-282).



At the 0.01 level of confidence there were no significant differences between means for the three embodiments at any grade level. The hypothesis is therefore accepted for all three grades.

TABLE 10

COMPARISON OF EMBODIMENT MEANS FOR EACH GRADE  
AND THE TOTAL SAMPLE

| Subjects     | Embodiment means |       |      | F    | d.f.  | Conservative<br>probability<br>of F |
|--------------|------------------|-------|------|------|-------|-------------------------------------|
|              | Spinner          | Block | Box  |      |       |                                     |
| Total sample | 7.61             | 7.35  | 7.81 | 2.23 | 1, 71 | 0.140                               |
| Grade 1      | 6.46             | 6.13  | 7.29 | 6.57 | 1, 23 | 0.017                               |
| Grade 2      | 6.96             | 7.29  | 7.42 | 0.73 | 1, 23 | 0.401                               |
| Grade 3      | 9.42             | 8.63  | 8.71 | 2.76 | 1, 23 | 0.110                               |

Hypothesis Six Restated: There are no significant main effects due to embodiment on the probability test performance when sex, grade, and IQ are used as blocking variables in pairs.

The hypothesis was tested using a three-way analysis of variance with the embodiment factor as a repeated measure. Three analyses were carried out using, two at a time, the factors sex, grade, and IQ for the blocking variables. A summary of the analyses of variance within subjects relating to embodiment is given in Table 11.

None of the F-ratios for embodiment (E) are significant at



the 0.01 level of confidence. The hypothesis, that there are no significant main effects due to embodiment, is retained.

TABLE 11

SUMMARY OF THREE ANALYSES OF VARIANCE WITHIN SUBJECTS  
DUE TO EMBODIMENT AND SEX, GRADE, AND IQ

| Source of variance | df  | MS   | F    | Prob. of F |
|--------------------|-----|------|------|------------|
| (a) Embodiment (E) | 2   | 3.81 | 2.40 | 0.095      |
| E x Sex (A)        | 2   | 1.25 | 0.79 | 0.456      |
| E x Grade (B)      | 4   | 5.37 | 3.38 | 0.011      |
| E x A x B          | 4   | 2.16 | 1.36 | 0.250      |
| Error              | 132 | 1.59 |      |            |
| (b) Embodiment (E) | 2   | 3.81 | 2.26 | 0.108      |
| E x Sex (A)        | 2   | 1.25 | 0.74 | 0.477      |
| E x IQ (C)         | 2   | 4.85 | 2.87 | 0.060      |
| E x A x C          | 2   | 0.37 | 0.22 | 0.806      |
| Error              | 136 | 1.69 |      |            |
| (c) Embodiment (E) | 2   | 3.81 | 2.48 | 0.088      |
| E x Grade (B)      | 4   | 5.37 | 3.50 | 0.010      |
| E x IQ (C)         | 2   | 4.85 | 3.15 | 0.046      |
| E x B x C          | 4   | 2.08 | 1.35 | 0.254      |
| Error              | 132 | 1.54 |      |            |



Hypothesis Seven Restated: There are no significant interactions between the embodiment and the factors sex, grade, and IQ.

The same three-way analysis of variance used to test hypothesis six also tested for interactions between the variables. The data in Table 11 indicate that the only interaction significant at the 0.01 level is between grade level and embodiment when grade and IQ are the blocking variables. When considered with sex, grade just fails to show a significant interaction with embodiment at the 0.01 level of confidence.

The hypothesis of no interaction with embodiment is accepted for sex and IQ but judgment is reserved in the case of grade.

#### IV. THE EFFECTS OF SEX, GRADE, AND IQ

Hypothesis Eight Restated: There is no significant effect on performance on the probability test due to (a) sex, (b) grade, and (c) IQ.

This hypothesis was tested by a three-way analysis of variance for the three criterion variables, total score, concept subtest score, and quantitative subtest score. The results are reported separately for each criterion in Table 12.

On the total test score, all three factors are shown to have an effect which is significant at the 0.01 level of confidence. On the concept subscore, the effects of grade and IQ are shown to be significant at the 0.01 level, but sex only at the 0.05 level. On the quantitative subscore all factors appear to have a significant effect only at the 0.05 level of confidence.





TABLE 12

SUMMARY OF ANALYSIS OF VARIANCE  
FOR RESPONSES TO THE TEST AND SUBTESTS

| Source    | df | Total |       | Concept |       | Quantification |       |
|-----------|----|-------|-------|---------|-------|----------------|-------|
|           |    | F     | Prob. | F       | Prob. | F              | Prob. |
| Sex (A)   | 1  | 9.02  | 0.004 | 4.55    | 0.037 | 5.40           | 0.024 |
| Grade (B) | 2  | 13.67 | 0.000 | 16.63   | 0.000 | 3.29           | 0.044 |
| IQ (C)    | 1  | 13.13 | 0.001 | 11.69   | 0.001 | 4.84           | 0.032 |
| A x B     | 2  | 1.10  | 0.340 | 0.23    | 0.795 | 1.56           | 0.219 |
| B x C     | 2  | 1.96  | 0.149 | 1.23    | 0.300 | 1.49           | 0.233 |
| A x C     | 1  | 0.00  | 0.985 | 2.48    | 0.120 | 1.42           | 0.237 |
| A x B x C | 2  | 0.17  | 0.845 | 0.99    | 0.378 | 0.25           | 0.780 |
| Error     | 60 |       |       |         |       |                |       |

On the basis of the analysis reported above, the hypothesis is rejected for each of the factors on the total score. It did make a difference to the total test score whether the subject was a boy or a girl, whether in grade one, two, or three, and whether high or low in general reasoning ability.

These differences in performance are evident in Table 13 which gives the mean on each criterion measure for each grouping of subjects, and in Table 14, which gives the means for sex and IQ groupings nested in grades. Figures 2, 3, and 4 are graphical representations of the means in Table 14 for the total score on the probability test and the two subscores. On each criterion



TABLE 13  
CRITERION MEANS BY SEX, GRADE, AND IQ

| Criterion<br>measure | Factors |      |       |      |      | Grand<br>mean | S.D. |      |      |
|----------------------|---------|------|-------|------|------|---------------|------|------|------|
|                      | Sex     |      | Grade |      |      |               |      | IQ   |      |
|                      |         |      | 1     | 2    | 3    |               |      |      |      |
|                      | B       | G    |       |      |      |               |      | Low  | High |
| Total (/39)          | 21.4    | 24.2 | 19.9  | 21.8 | 26.8 | 20.8          | 24.8 | 22.8 | 5.9  |
| Concept (/21)        | 16.6    | 17.8 | 15.1  | 17.1 | 19.4 | 16.3          | 18.2 | 17.2 | 3.3  |
| Quant. (/18)         | 4.8     | 6.5  | 4.8   | 4.7  | 7.3  | 4.6           | 6.7  | 5.6  | 4.1  |



TABLE 14

CRITERION MEANS BY SEX AND IQ NESTED WITHIN GRADES

| Criterion<br>measure | 1     |      |      | 2    |      |      | 3    |      |      |
|----------------------|-------|------|------|------|------|------|------|------|------|
|                      | Grade |      |      |      |      |      |      |      |      |
|                      | Sex   |      |      | G    |      |      | G    |      |      |
| IQ                   | B     |      |      | B    |      |      | B    |      |      |
|                      | L     | H    | L    | L    | H    | L    | L    | H    | H    |
| Total (/39)          | 14.5  | 21.4 | 18.1 | 18.7 | 23.8 | 20.6 | 23.9 | 24.0 | 30.4 |
| Concept (/21)        | 12.5  | 15.5 | 14.9 | 14.0 | 19.3 | 17.1 | 18.3 | 19.4 | 20.3 |
| Quant. (/18)         | 2.0   | 5.9  | 3.3  | 4.7  | 4.5  | 3.4  | 5.6  | 4.6  | 10.1 |





measure girls outscored boys, grade three scores were higher than those of grade one or grade two subjects, and the high-IQ subjects scored higher than the low-IQ subjects.

Hypothesis Nine Restated: There are no significant interactions between the independent variables sex, grade, and IQ and the criterion measures in the test.

The analysis of variance summarized in Table 12 included tests for interaction effects. No significant interactions were found between the variables. The lowest probability assigned to any of the interactions is 0.120 for the Sex x IQ interaction. The hypothesis is therefore accepted. Sex, grade, and IQ did not interact with one another to produce any differential effects on probability test achievement.

FIGURE 2

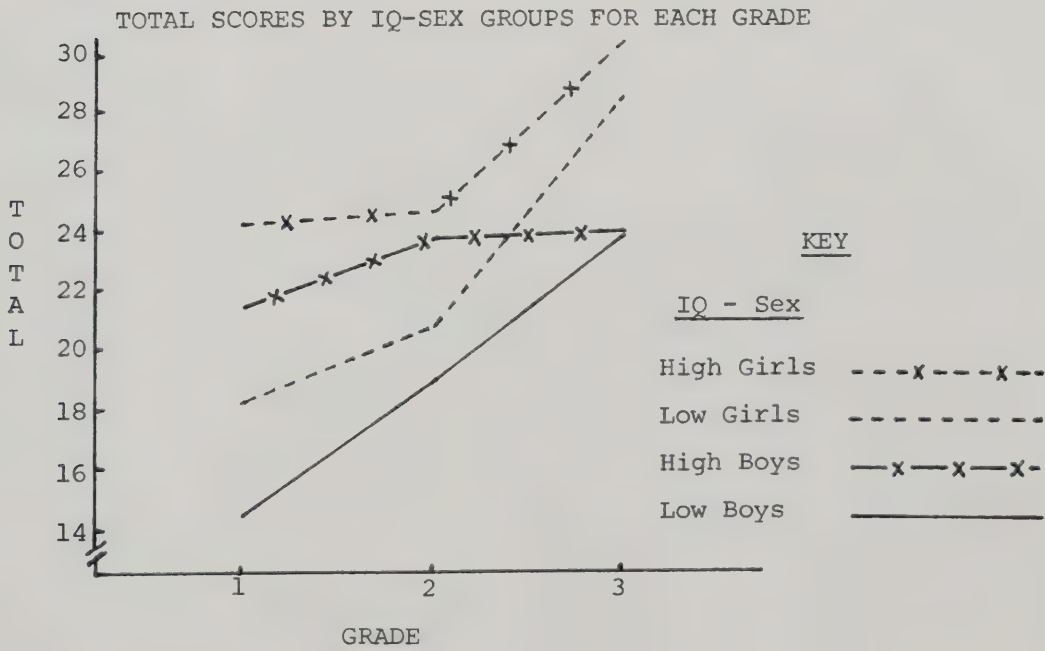




FIGURE 3

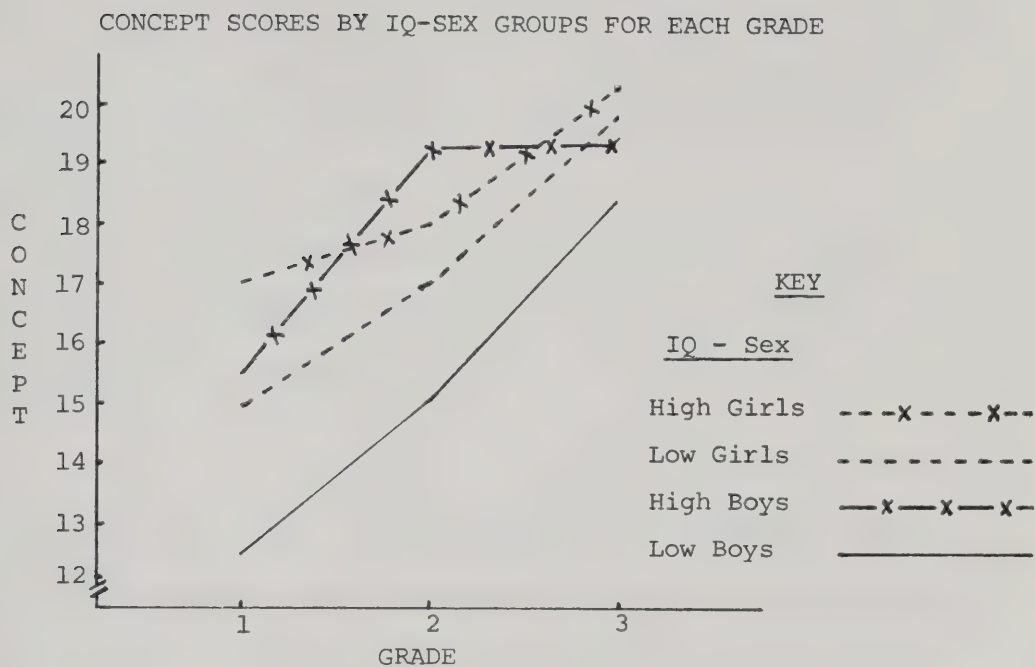
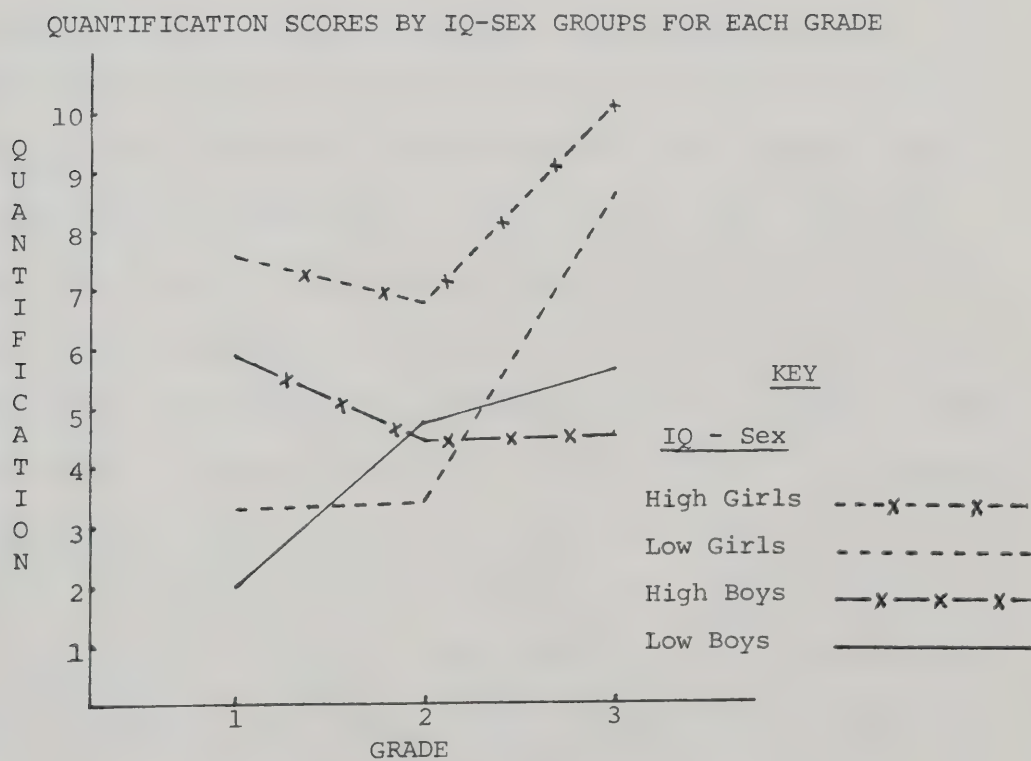


FIGURE 4





## V. RATIONALIZATION USED BY SUBJECTS

In addition to the foregoing questions and hypotheses, a subsequent question, arising from the third purpose of the probability test, was asked:

What rationalizations are given by subjects for their responses to probability questions? Are there any observable trends related to grade level?

As indicated in chapter 3, subjects were found to rationalize their responses in one of three ways: (a) by correct reasoning, (b) by reference to color preference, position, or quasi-magical properties, or (c) by indicating they had no reason or didn't know.

These categories applied only to responses related to concepts. When asked for reasons for their answers to the quantification questions, only two subjects attempted to explain their answers. One was a second-grade girl who attended closely to the relative frequencies of the favorable color. The other was a third-grade boy who was reluctant to make predictions about events which he perceived to occur with random and uncertain frequency. For example, when asked to predict the number of times that the 2-3-1 block would turn up red out of six rolls he replied "You couldn't tell really, as it could be red every time in one lot of throws and not at all in another lot". When pressed for a commitment he said "Three times" and gave as his reason "Three faces are red". From these and other comments, the experimenter judged this subject to be well in advance of his peers in



understanding probability concepts.

The main analysis relating to the subsequent question, then, concerns rationalizations for concept responses. Table 15 contains frequencies of the types of rationalizations given for responses to items about all concepts except the first one, events in a sample space. As outlined in chapter 3, subjects were not required to explain the basis for their responses to items one, two, and three of the test.

In a number of cases a subject gave a correct rationalization even though his actual response may have been incorrect. This was particularly so with items relating to the fifth concept and accounts for an apparent discrepancy between the percentages given for this concept in Tables 6 and 15.

A number of trends are apparent. For four of the five concepts an acceptable rationalization was given more than 65% of the time with the whole sample. Except for the fifth concept, grade three subjects rationalized correctly more than 95% of the time. Color preference, position, and quasi-magical reasons seemed to be evident only in grade one and two subjects (with one exception) and then only regarding concept two to any marked extent. Concept five proved to be the most difficult for all students to rationalize their responses.

#### Examples of Rationalization

The following are some examples of rationalizations used by the subjects. These are grouped according to the concepts for which they were given and are referred to as types (a) or (b).





TABLE 15

## FREQUENCIES OF RATIONALIZATION RESPONSES FOR CONCEPTS

BY GRADES AND FOR THE TOTAL SAMPLE

| Concept                                 | Type of<br>reason given <sup>1</sup> | Grade (N=24) |    |    | Total sample<br>(N=72) |    |
|-----------------------------------------|--------------------------------------|--------------|----|----|------------------------|----|
|                                         |                                      | 1            | 2  | 3  | n                      | %  |
| 2. Most<br>favorable<br>event           | a                                    | 12           | 13 | 23 | 48                     | 67 |
|                                         | b                                    | 10           | 8  | 1  | 19                     | 26 |
|                                         | c                                    | 2            | 3  | 0  | 5                      | 7  |
| 3. Most<br>favorable<br>sample space    | a                                    | 18           | 23 | 24 | 65                     | 90 |
|                                         | b                                    | 2            | 1  | 0  | 3                      | 4  |
|                                         | c                                    | 4            | 0  | 0  | 4                      | 6  |
| 4. Equally<br>favorable<br>sample space | a                                    | 14           | 15 | 24 | 53                     | 74 |
|                                         | b                                    | 1            | 1  | 0  | 2                      | 3  |
|                                         | c                                    | 9            | 8  | 0  | 17                     | 23 |
| 5. Impossible<br>event                  | a                                    | 10           | 12 | 18 | 40                     | 56 |
|                                         | b                                    | 0            | 0  | 0  | 0                      | 0  |
|                                         | c                                    | 14           | 12 | 6  | 32                     | 44 |
| 6. Certain<br>event                     | a                                    | 19           | 20 | 23 | 62                     | 86 |
|                                         | b                                    | 1            | 0  | 0  | 1                      | 1  |
|                                         | c                                    | 4            | 4  | 1  | 9                      | 13 |

1. a - correct; b - incorrect; c - none or didn't know



Most favorable event.

(a) Subjects most often responded that there was more of the selected color, or they counted segments, faces, or counters and concluded that there were numerically more of them of that color than other colors.

(b) Those responses categorized as type (b) included the following reasons, each of which indicates what the subject appeared to be attending to: "I like red; It's my favorite color", "Blue will win because it is my second favorite color", and "It's my (or the) best color".

Other reasons which were not related to a color preference included "It's the first one" (from a subject who chose the device nearest to him), "It's faster" (from one subject who attributed this trait to the color red), and "It's lighter" (from another subject who thus explained why the yellow side of the block would occur most often).

Most favorable sample space.

(a) Subjects tended to either give the correct rationalization for this concept or give no reason at all. Examples were: "Each color has two", "There are two of each", and "The colors are all even (or equal)". One subject incorrectly selected the 2-3-1 device and reasoned "It has all the colors". He appeared not to see the importance of the relative frequencies of the possible events.



### Impossible event.

(a) An example of correct reasoning was "There's no yellow, it (the yellow marker) wouldn't get to move at all,"

(b) Examples of false reasoning were "I don't like yellow, it always loses" and "Blue couldn't win (selecting the 1-2-3 device) as there's only one blue". In one case a subject selected the 4-1-1 device and explained "One color can win and two can lose".

### Certain event.

(a) Most subjects who gave reasons for their responses to items related to this concept were correct in their rationalizations, and also in their responses. A typical response was "It's all red, so only red would go. It would always win".

(b) The one erroneous reason given was related to color preference which caused the particular subject to select the 1-4-1 device and explain "Red would always win because it's my favorite color".

### Concluding Statement

Although there is often considerable interest in "erroneous" rationalizations used by children, and indeed much can be learned from them, the researcher was impressed by the quality and the frequency of subjects' correct rationalizations on most of the concept items in the probability test. On the average, 75% of rationalizations were correct in the whole sample, while over 93% of the grade three responses were correct.



## VI. SUMMARY

Chapter IV contains the results of answering three questions and testing seven hypotheses associated with the four major purposes of the present study.

The first purpose of the study was to determine what percentage of students appeared to have an understanding of the six basic probability concepts. It was found that over 74% of subjects in the whole sample showed an understanding of five of the six concepts. Over 92% of the grade three subjects showed such understanding and there was a general improvement in performance evident as age increased. The one concept that received fewest correct responses in all grades was number five, impossible event. The next poorest response was on concept number four, equally favorable sample space, which was correctly identified by 74% of all the subjects, but only by 58% of the grade one children.

When the rate of response was compared to what might be expected from purely random responses, it was found that the responses on all six concepts were significantly better than chance at the 0.01 level of confidence.

The second purpose of the study dealt only with quantification of probability. Scores were much lower than on the concept items; an average of only 42% of subjects responded correctly on the quantification items. Grade three subjects performed significantly better than grade one or two subjects. No difference was found between grade one and two scores on either subtest.





There was no significant difference between scores on three different probability settings (2-2-2, 3-2-1, and 4-1-1) but as the number of trials in the sample space increased the number of correct predictions decreased.

The third purpose of the study was to examine the effect of embodiment on pupils' responses to probability questions. No significant main effects were found due to embodiment at any of the grade levels. No interactions were evident between embodiment and sex or IQ but judgment was reserved in the case of interaction between embodiment and grade.

The fourth purpose of the study was to investigate the effects of sex, grade, and IQ on the criterion scores. Significant effects were found, at the 0.01 level of confidence, for all three factors. Girls were found to score higher than boys, grade three subjects scored higher than those in grades one and two, and the high-IQ group scored higher than the low-IQ group. There were no interactions between sex, grade, and IQ on probability test achievement.

The fifth discussion in chapter 4 related to the rationalizations used by subjects to explain their responses. Only 3% of subjects offered reasons for quantification responses while over 85% of subjects responded on four of the six concepts, and 77% and 56% on the other two. Of the rationalizations given, an average of 75% were correct across the whole sample, while 93% of the grade three rationalizations were correct.

In the next and final chapter, the study is summarized in terms of its purposes, the instrumentation and procedures used,



and the analysis of the results. Implications for teaching, curriculum design, and further research are made in conclusion.



## CHAPTER 5

## SUMMARY, DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS

## I. SUMMARY OF THE INVESTIGATION

The present study developed from a need for more information about young children's understanding of basic probability notions. It is widely accepted that the elementary school curriculum should include topics on probability in order to give children early experience in dealing with degrees of uncertainty and to prepare them for later studies in statistics. According to Ausubel (1968), meaningful learning experiences can only be constructed on the basis of the learner's existing knowledge and understanding. The first task for a curriculum writer or a teacher is to ascertain the level of readiness a learner has for a topic.

The main purpose of the present study was to investigate young children's readiness for instruction in probability by determining how well six basic concepts were understood by children in grades one, two, and three. A sample of 72 pupils was chosen, 24 in each grade, and a test was administered to each one in an interview with the researcher. A game was employed in the test items as a motivating agent and to encourage subjects to maximize their responses. There were 21 questions dealing with the six concepts. In addition 18 quantification questions were added and subjects were asked to give



rationalizations for each response. The materials and apparatus used in the study were specially constructed so as to represent each of five probability settings in three embodiments, spinner, block, and box. This allowed for testing for the effect of embodiment on subjects' responses.

A test-retest reliability check on the probability test gave Pearson product-moment correlation coefficients of 0.837, 0.796, and 0.753 for total and subtest scores; all were significantly different from zero at the 0.01 level.

Analysis of the results was organized into five sections corresponding to the four purposes of the study and a report on the rationalizations used by subjects. Three questions were asked and seven hypotheses were tested. A brief summary of the main findings is now given under five headings.

### Concepts

Four of the six concepts tested were understood by at least 75% of the subjects in each grade. They were: sample space events, most favorable event, most favorable sample space, and certain event. The other two concepts, equally favorable sample space and impossible event, were understood by 74% and 50% respectively of the total sample. The scores on all concept items were significantly greater than was attributable to chance.

### Quantification

The quantification of probability proved to be less understood than the concepts. The average correct response rate for all subjects on the 18 quantification items was 42%.





There was no significant difference in scores due to different probability settings but performance decreased as the number of trials in the sample space increased from six to twelve.

### Embodiment

No significant effect was found due to embodiment at any grade level, and no interactions were evident between embodiment and sex or IQ.

### Significant Factors

Sex, grade, and IQ were all found to have significant effects on the total test score. Girls scored higher than boys, grade three subjects scored higher than those in grades one and two, and the high-IQ group scored higher than the low-IQ group. Only grade and IQ had significant effects on the concept score and no significant effects were found on the quantification score. No interactions were found between sex, grade, and IQ on any of the criterion measures.

### Rationalization

When subjects were asked for their answers to the test questions, one of three types of responses was given: (a) a correct explanation based on the proportions within the probability settings in question, (b) an erroneous explanation involving the subject's color preference, the position of the devices, or quasi-magical properties seen in the situation, or (c) no answer, or an indication of having no reason or of not knowing.



Considering all of the concept items, an average of 80% of subjects rationalized with type (a) or type (b) responses. Of these, 75% were correct for the whole sample and 93% were correct for grade three. Samples of rationalizations given for the responses for each concept were included with the report in chapter 4. Only two subjects gave rationalizations for quantification responses. Most gave no reason for their predictions on these items.

## II. DISCUSSION OF THE FINDINGS

### Concepts

The first concept, events in a sample space, presented little problem for any of the children. The researcher found that most of the subjects were almost puzzled that such an "obvious" question be asked as, for example, "What color could I get if I spin this spinner?" Once the meaning of the question was grasped (and that sometimes proved to be the main problem) almost all subjects readily supplied the correct answer. The children in the study were able to recognize all the possibilities in a situation when they understood that this was what was required of them.

The second concept, the most favorable event, was understood by 83% of all the subjects and by 96% of the third graders. Most of the children simply selected from the three colors present the color that was most plentiful. This meant counting the number of sectors on the spinner, the faces on the block, or the counters



in the box that were of each color. Eleven of the twelve subjects who responded incorrectly were in grades one and two and selected their favorite color or the closest one to them, or chose on the basis of imputed advantage of one color over the others.

The third concept, most favorable sample space, was understood by almost all the second and third graders and 75% of first graders. In the items relating to this concept subjects were presented with three probability settings, for example, 2-2-2, 3-2-1, and 4-1-1. It appeared to be relatively easy for most subjects to compare these settings and to choose the one which maximized the chance of a particular color being the outcome, 4-1-1 and BLUE in the example. It made little difference whether the settings were embodied by spinners, blocks, or boxes. The main rationalization given by subjects concerned the relative amount of the favored color. With each embodiment subjects were able to count discrete units of each color, make comparisons of the amount of color, and make their choice on that basis. In each case their reasoning was correct, there being an equal number of units (six) in each setting and device. Subjects in grade three often verbalized their reasoning in clearer terms than the researcher had expected.

The fourth concept, equally favorable sample space, was the second most difficult for the subjects with an overall correct response rate of 74%. Third-grade subjects (92% correct) had little difficulty selecting the correct 2-2-2 setting, but many of the grade one and two children balked at the three 3-way



comparisons that confronted them. For some of these children the exact meaning of "fair", "same chance", or "equal chance" may not have been clear when there were three possible outcomes. This was indicated by several subjects selecting two of the devices which had the same amount of one particular color. These responses serve to emphasize the need for extreme care when using verbal instructions or questions with 6- and 7-year olds. The same problem was not encountered with the third-grade subjects.

The fifth concept, impossible event, was by far the most difficult for all subjects. The grade one response was no better than a chance response and only 67% of grade three subjects responded correctly. The most common incorrect response was the selection of the 3-2-1 or the 4-1-1 setting (or equivalent) and identification of the one-unit color as the impossible outcome. To ensure that the subjects giving these responses were not just confusing "impossible" with "unlikely", the investigator always repeated the question with emphasis on "never win" and always lose". For more than half of the grade one and two subjects it made little difference; the two words, impossible and unlikely, appeared to suggest the same idea. A third of the grade three pupils also saw no differences.

Several subjects selected the 6-0-0 setting in their response and nominated as impossible outcomes the colors with no representation in the setting. The investigator accepted such a response as indicative of an understanding of the concept in





question. This judgment was generally confirmed by later asking the subject what would happen in the game if the 3-3-0 device were used. Except for one case (no response), the subjects responded in terms of one event never occurring or one color (or marker) never winning.

The final concept, certain event, was correctly understood by 85% of all subjects, and by 96% of the third graders. The responses were usually given quickly and subjects often reacted as though the questions relating to this concept were "obvious", as with the first concept. Few of the subjects had difficulty understanding the meaning of the questions. The incorrect responses generally came from some of the same subjects who were wrong on the concept five questions, and for similar reasons. For example, the 3-2-1 or 4-1-1 setting was selected and BLUE identified as the certain outcome. The ideas of certainty and likelihood were apparently being equated by these students; this tendency was not shown by any of the third grade subjects.

### Quantification

Subjects' responses on the quantification questions indicated a low level of understanding of the relationship between the proportions in the settings and the chances of the various outcomes. The numbers of correct responses made by grades one and two were not significantly better than chance. The significantly higher scores by grade three subjects, even though still only 50% were correct, suggest that grade three is the earliest that such quantification ideas should be introduced.



There was little difference in the number of correct responses due to the probability setting, but doubling the number of trials (from six to twelve) resulted in more than a 50% drop in correct responses.

### Embodiment

The probability test was arranged so that the subjects answered thirteen questions in each of the three embodiments, spinner, block, and box. No significant differences at the 0.01 level were found between the mean scores for these embodiments for any of the grades. This was decided using the conservative probability of F in a univariate analysis with embodiment as a repeated measure. Using a three-way analysis with sex, grade, and IQ as blocking variables in pairs, no significant interactions were found between embodiment and sex or IQ. Judgment was reserved in the case of interaction with grade as it was significant when blocked with IQ, but not significant when blocked with sex. (Both probabilities of F were close to 0.01.) This meant that subjects' responses to embodiment tended to vary according to their grade level. This variation is shown in Table 16 by the rank of embodiment responses within grades. The grade one subjects responded correctly to the box embodiment 17% more often than to the block, and 13% more often than to the spinner. The grade three subjects, on the other hand, responded correctly to the spinner embodiment 9% more often than to the block, and 8% more often than to the box embodiment. The grade two means were in the order box, block, and spinner within a 6% range.



TABLE 16

## RANK OF EMBODIMENT RESPONSES WITHIN GRADES

| Embodiment | Grade 1 | Grade 2 | Grade 3 |
|------------|---------|---------|---------|
| Spinner    | 2       | 3       | 1       |
| Block      | 3       | 2       | 3       |
| Box        | 1       | 1       | 2       |

A possible reason for the grade one preference for the box embodiment is the easier task of counting discrete objects rather than proportions on a disc or faces on a cube. Many of the early number activities at school involve counters; this would predispose younger children to this mode of counting. By third grade, subjects had gained greater ability to count and reason about non-discrete items that were fixed in their spatial relationship.

#### Significant Factors

The findings of the present study agree with those of earlier studies that grade (or age) and IQ are significant factors in responses to probability questions. There was a highly significant difference ( $p < 0.001$ ) between the grade three performance and grades one and two on the concept and total scores, but little difference between grades one and two on any of the criterion scores. This tends to indicate a substantial increase in understanding of probability concepts in children as they pass through grade three, about age 8 years. According



to Piaget's theory, this jump in understanding is the beginning of stage II development, proposed by Piaget to begin at around seven years.

In the present investigation, sex was a significant factor in the total test score but not in either of the subscores. Girls scored higher than boys on all criterion measures and within most IQ and grade groupings.

On the quantification subtest, none of the factors was significant at the 0.01 level. The subjects in all three grades, of both sexes, and in both IQ groups found the quantification questions uniformly difficult, although grade three scores were higher than grades one and two, as already mentioned.

#### Rationalizations

The main impression formed by the investigator as a result of the subjects' rationalizations is that most of the children in the sample correctly understood at least four of the basic concepts and could express quite adequately the basis of their responses to the probability concept questions. The absence of rationalizations for quantification responses indicated that many subjects had little understanding of the numerical relationships between the probability settings and the frequencies of the outcomes. Many responses appeared to be guesses, although third-grade subjects scored well on the six-trial questions and may have been at the threshold of understanding quantification in sample spaces with a small number of events.





### III. SOME IMPLICATIONS OF THE FINDINGS

#### For Teachers

From the discussion in the previous section a number of classroom implications arise. The first relates to the main finding that at least four of the probability concepts were understood by the majority of the grade one, two, and three subjects. The writer would encourage teachers to provide experiences in probability in these grades (especially grade three), experiences which allow children to become involved in the real world activities of decision making. This is not too ambitious for young children, for their vocabulary, and performance in this study, indicate they already have begun to appreciate the uncertainty in many situations which they face. Having begun to accumulate ideas about chance events at a young age, children can easily form incorrect intuitions which are difficult to alter, as Wilkinson and Nelson (1966) found. As teachers, we need to include probability experiences in the lower grades that will produce correct intuitions and so build a better base for the more formal study of probability and statistics in later grades.

Further suggestions can be made from the writer's observations in this study:

1. Beginning activities should involve only sample spaces with small numbers of outcomes which are easily identified by the students as possible occurrences.
2. Comparison and choice responses should be used initially



in preference to prediction responses. The probability settings being compared should have the same number of units to allow comparisons on the basis of absolute number. Otherwise ratios with unequal denominators are involved; these are not handled with any substantial skill until later grades.

3. Care should be taken to minimize misunderstanding due to word meaning. The problem of communicating ideas accurately is ever present at all grade levels. As far as possible, non-verbal methods and materials that communicate ideas and command action or decision with little direction needed from the teacher should be used with young children.

#### For the Curriculum

The present study indicates that grade one, two, and three children have sufficient intuitive understanding of probability concepts to form a basis for the development of special instructional units appropriate to their grade level. Early activities in grade one and two need to provide informal consolidation of existing concepts and ideas. From grade three onwards, units should provide a wider range of experiences and activities in which students are led to further concepts and into quantification of probability.

#### IV. RECOMMENDATIONS FOR FURTHER RESEARCH

No attempt was made in the present study to test the feasibility of teaching probability topics to grades one, two, and three. The present study was designed mainly to survey the



field for understanding of six basic concepts. There are two main recommendations for further research: a replication of the study, with improvements, and the development and trial of appropriate instructional units in probability for primary grades.

One improvement in a replication study would be to use a larger sample chosen from a wider variety of schools. Changes could be made in the instrument to include the testing of a further concept, unlikely event, for comparison with the response to impossible event. Special attention would need to be given to the wording and presentation of these items. Another suggestion would be to use grade two, three, and four students as little difference was found between grades one and two. There may then be little difference between grades three and four as third-graders in the present study scored near the maximum on the concept subtest. Such a replication may, on the other hand, give further insight into quantification understanding around that age or grade level.

Many worthwhile studies have been done regarding instructional units and methods appropriate to the senior grades in the elementary school. The implications stated in section III of this chapter lead to a recommendation for comparable research and development at the junior grade level. Wilkinson and Nelson (1966) gave six practical suggestions to those who would design probability units for elementary school children. Shepler (1969) proposed a sequence for developing research-based curriculum materials using behavioral objectives and task analysis. Using



a combination of these guidelines (outlined in chapter 2 of this report), and taking the four concepts sample space, most favorable event, most favorable sample space, and certain event as a basis, a unit on probability should be designed and tested with junior grades.

Teachers-in-training need to be instructed in recent curriculum changes. Teacher-preparation programs and in-service courses need to include updated components related to the teaching of probability in elementary grades if any significant implementation of the recommended changes in curriculum is to occur.





## BIBLIOGRAPHY



## BIBLIOGRAPHY

- Ausubel, D. P. Educational Psychology: A Cognitive View. New York: Holt, Rinehart, & Winston, 1968.
- Blishen, B. R., Jones, F. E., Naegele, K. D., & Porter, J. Canadian Society: Sociological Perspectives. Totonto: MacMillan, 1968.
- Brackbill, Y., Kappy, M. S., & Starr, R. H. Magnitude of Reward and Probability Learning. Journal of Experimental Psychology, 1962, 63, 32-35.
- Bruner, J. S. "On the Learning of Mathematics - A Process Orientation". In D. B. Aichele & R. E. Keys (Eds.), Readings in Secondary School Mathematics. Boston: Prindle, Weber, & Schmidt, 1974.
- Cambridge Conference on School Mathematics. Goals for School Mathematics. N. Y.: Houghton Mifflin, 1963.
- Carlson, J. S. Children's Probability Judgments as Related to Age, Intelligence, Socio-Economic Level and Sex. Human Development, 1969, 12(3), 192-203.
- Cohen, J. Subjective Probability. Scientific American, Nov. 1957, 197, 128-138.
- Davies, C. M. Development of the Probability Concept in Children. Child Development, 1965, 36, 779-788.
- Doherty, Joan. Level of Four Concepts of Probability Possessed By Children of Fourth, Fifth, and Sixth Grade before Formal Instruction (Doctoral Dissertation, University of Missouri, 1965). Dissertation Abstracts, Dec. 1966, 27, 1703A. (University Microfilms No. 65-14, 465).
- Engel, Arthur. Mathematical Research and Instruction in Probability Theory, The Mathematics Teacher, Dec. 1966, 771-782.
- Engel, Arthur. Teaching Probability in the Intermediate Grades. International Journal of Mathematical Education in Science and Technology, 2, July/September, 1971, 243-294.
- Estes, W. K. Research and Theory on the Learning of Probabilities. Journal of the American Statistical Association, 1972, 67, 81-102.
- Estes, W. K. The Cognitive Side of Probability Learning, Psychological Review, 1976, 83(1), 37-64. (a)



- Estes, W. K. Some Functions of Memory in Probability Learning and Choice Behavior. In G. H. Bower (ed.), The Psychology of Learning and Motivation (Vol. 10). New York: Academic Press, 1976. (b)
- Ferguson, G. A. Statistical Analysis in Psychology and Education, 3rd Edition. N. Y.: McGraw-Hill, Inc., 1971.
- Fischbein, E. The Intuitive Sources of Probabilistic Thinking in Children. Dordrecht, Holland: D. Reidel Pub. Co., 1975.
- Fischbein, E. Probabilistic Thinking in Children and Adolescents. In Forschung zum Prozeß des Mathematiklernens (Addresses Given at the Third International Congress on Mathematics Education at Karlsruhe). Universität Bielsfeld, 1976, 23-42.
- Fischbein, E., Pampu, I., & Manzat, I. Comparison of Ratios and the Chance Concept in Children. Child Development, June 1970, 41, 377-389.
- Flavell, J. H. The Developmental Psychology of Jean Piaget. New York: D. Van Nostrand Co. Inc., 1963.
- Gipson, J. Teaching Probability in the Elementary School: An Exploratory Study (Doctoral Dissertation, University of Illinois, 1971). Dissertation Abstracts International, Feb. 1972, 32, 4325A-4326A. (University Microfilms No. 72-6933, 374).
- Glass, G. V., & Stanley, J. C. Statistical Methods in Education and Psychology. Englewood Cliffs, N. J.: Prentice-Hall, 1970.
- Goldberg, S. Probability Judgements by Pre School Children: Task Conditions and Performance. Child Development, March 1966, 37, 157-167.
- Harvey, J. G. An Immodest Proposal for Teaching Statistics and Probability. School Science and Mathematics, March 1972, 129-144.
- Hoemann, H. W. & Ross, B. M. Children's Understanding of Probability Concepts. Child Development, 1971, 42, 221-236.
- Jones, G. A. The performances of First, Second, and Third Grade Children on Five Concepts of Probability and the Effects of Grade, IQ, and Embodiments on Their Performance (Doctoral Dissertation, University of Indiana, 1974). Dissertation Abstracts International, Jan. 1975, 35, 4272A-4273A. (University Microfilms No. 75-1710, 387).



- Leffin, W. W. A Study of Three Concepts of Probability Possessed by Children in the Fourth, Fifth, Sixth, and Seventh Grades (Doctoral Dissertation, University of Wisconsin, 1968). Dissertation Abstracts, June 1969, 29, 4188A-4189A. University Microfilms No. 68-17, 913).
- Lovell, K. Proportionality and Probability. In NCTM, Piagetian Cognitive-Development Research and Mathematical Education. Washington, D. C.: NCTM, 1971.
- McLeod, G. K. An Experiment in the Teaching of Selected Concepts Of Probability to Elementary School Children (Doctoral Dissertation, Stanford University, 1971). Dissertation Abstracts International, Sept. 1971, 32, 1359A. (University Microfilms No. 71-23, 535, 231).
- Messick, S. J., & Solley, C. M. Probability Learning in Children: Some Exploratory Studies. Journal of Genetic Psychology, 1957, 90, 23-32.
- Mullenex, J. L. A Study of the Understanding of Probability Concepts by Selected Elementary School Children (Doctoral Dissertation, University of Virginia, 1968). Dissertation Abstracts, May 1969, 29, 3920A. (University Microfilms No. 69-4010, 89).
- National Advisory Committee on Mathematical Education. A Technical Report: Overview and Analysis of School Mathematics Grade K - 12. Washington, D. C.: Conference Board of the Mathematical Sciences, 1975.
- Offenbach, Stuart I. Studies of Children's Probability Learning Behavior: I. Effect of Reward and Punishment at Two Age Levels. Child Development, 1964, 35, 709-715.
- Offenbach, Stuart I. Studies of Children's Probability Learning Behavior: II. Effect of Method and Event Frequency at Two Age Levels. Child Development, 1965, 36, 951-962.
- Ojemann, R. H., Maxey, J. E., & Snider, B. C. The Effect of a Program of Guided Learning Experiences in Developing Probability Concepts at the Third-Grade Level. Journal of Experimental Education, Summer 1965, 33(4), 321-330.
- Ojemann, R. H., Maxey, J. E., & Snider, B. C. Further Study of Guided Learning Experiences in Developing Probability Concepts in Grade Three. Perceptual and Motor Skills, 1966, 23, 97-98.
- Piaget, Jean. The Psychology of Intelligence. Patterson, N. J.: Littlefield, Adams & Co., 1960.





- Piaget, J. & Inhelder, B. The Origin of the Idea of Chance in Children. New York: W. W. Norton & Co., Inc., 1975.
- Racha-Intra, S. Basic Inferential Statistics in Grade Nine. Unpublished Doctoral Dissertation, University of Alberta, Edmonton, 1977.
- Restle, F. Psychology of Judgment and Choice: A Theoretical Essay. New York: Wiley, 1961.
- Romberg, T. A. & Shepler, J. Retention of Probability Concepts: A Pilot Study into the Effects of Mastery Learning with Sixth-Grade Students. Journal of Research into Mathematics Education, Jan. 1973, 4(1), 27-32.
- S.M.S.G. Probability for Primary Grades, Teacher's Commentary. Pasadena, Ca.: A. C. Vroman, 1966.
- Shepler, J. L. Parts of a Systems Approach to the Development of a Unit in Probability and Statistics for the Elementary School. Journal for Research in Mathematics Education, Nov. 1970, 1(5), 197-205.
- Siegel, S. & Andrews, J. M. Magnitude of Reinforcement and Choice Behavior in Children. Journal of Experimental Psychology, 1962, 63, 337-341.
- Stevenson, H. W. & Weir, M. W. Variables Affecting Children's Performance in a Probability Learning Task. Journal of Experimental Psychology, 1959, 57, 403-412.
- Stevenson, H. W. & Weir, M. W. The Role of Age and Verbalization in Probability Learning. American Journal of Psychology, 1963, 76, 299-305.
- Stevenson, H. W., & Zigler, E. F. Probability Learning in Children. Journal of Experimental Psychology, 1956, 56, 185-192.
- Strohner, H. & Nelson, K. E. The Young Child's Development of Sentence Comprehension, Influence of Event Probability, Nonverbal Context, Syntactic Form and Strategy. Child Development, 1974, 45, 567-576.
- Thorndike, R. L., Hagen, E., Lorge, I., & Wright, E. N. Canadian Cognitive Abilities Test, P2/F1; Examiner's Manual. Toronto: Thomas Nelson, 1970.
- U. S. Army, Ordnance Corps. Tables of the Cumulative Binomial Probabilities. Washington: 1952.



Wilkinson, J. D. & Nelson, O. Probability and Statistics -  
Trial Teaching in Sixth Grade. Arithmetic Teacher, Feb.  
1966, 13, 100-106.

Winer, B. J. Statistical Principles in Experimental Design.  
N. Y.: McGraw-Hill, 1971.

Yee, Albert H. Mathematics, Probability and Decision-Making.  
The Arithmetic Teacher, May 1966, 13, 385-387.

Yost, P. A., Siegel, A. E. & Andrews, J. McM. Non Verbal  
Probability Judgments by Young Children. Child Development,  
1962, 33, 769-780.



APPENDIX A

GAME BOARD USED IN PROBABILITY TEST



|        |        |        |        |
|--------|--------|--------|--------|
| FINISH | FINISH | FINISH | FINISH |
|        |        |        |        |
|        |        |        |        |
|        |        |        |        |
|        |        |        |        |
|        |        |        |        |
|        |        |        |        |
| START  | START  | START  | START  |





APPENDIX B

TABLES OF FREQUENCIES AND RELATIVE FREQUENCIES  
OF QUANTIFICATION RESPONSES



TABLE 17

FREQUENCIES AND RELATIVE FREQUENCIES OF RESPONSES ON  
THE SIX-TRIAL QUANTIFICATION QUESTIONS

| Subject's<br>response | Setting |       |       |       |       |       |
|-----------------------|---------|-------|-------|-------|-------|-------|
|                       | 2-2-2   |       | 3-2-1 |       | 4-1-1 |       |
|                       | f       | rf    | f     | rf    | f     | rf    |
| 1                     | 9       | 0.044 | 8     | 0.039 | 3     | 0.014 |
| 2                     | 94*     | 0.461 | 29    | 0.142 | 9     | 0.043 |
| 3                     | 40      | 0.196 | 93*   | 0.456 | 44    | 0.210 |
| 4                     | 26      | 0.127 | 25    | 0.123 | 90*   | 0.429 |
| 5                     | 13      | 0.064 | 19    | 0.093 | 34    | 0.162 |
| 6                     | 22      | 0.108 | 30    | 0.147 | 30    | 0.143 |

\* correct responses (also modal responses in these cases).



TABLE 18

FREQUENCIES AND RELATIVE FREQUENCIES OF RESPONSES  
ON THE TWELVE-TRIAL QUANTIFICATION QUESTIONS

| Subject's<br>response | Setting         |       |                  |       |                 |       |
|-----------------------|-----------------|-------|------------------|-------|-----------------|-------|
|                       | 2-2-2           |       | 3-2-1            |       | 4-1-1           |       |
|                       | f               | rf    | f                | rf    | f               | rf    |
| 1                     | 6               | 0.028 | 6                | 0.029 | 3               | 0.014 |
| 2                     | 34              | 0.156 | 7                | 0.034 | 5               | 0.024 |
| 3                     | 14              | 0.064 | 40               | 0.196 | 8               | 0.038 |
| 4                     | 31*             | 0.142 | 8                | 0.039 | 38              | 0.181 |
| 5                     | 18              | 0.083 | 13               | 0.064 | 8               | 0.038 |
| 6                     | 59 <sup>#</sup> | 0.271 | 60* <sup>#</sup> | 0.294 | 55 <sup>#</sup> | 0.262 |
| 7                     | 7               | 0.032 | 12               | 0.059 | 14              | 0.067 |
| 8                     | 13              | 0.060 | 14               | 0.069 | 29*             | 0.138 |
| 9                     | 5               | 0.023 | 7                | 0.034 | 7               | 0.033 |
| 10                    | 9               | 0.044 | 11               | 0.054 | 12              | 0.057 |
| 11                    | 5               | 0.023 | 2                | 0.010 | 9               | 0.043 |
| 12                    | 17              | 0.078 | 24               | 0.118 | 22              | 0.105 |

\* correct responses.

# modal responses.













**B30219**